AD-A254 278

ION PAGE

Form Approved
OMB No. 0704-0188

1. [illegible]	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 April 92	Final 1 April 89 - 31 March 92

4. TITLE AND SUBTITLE
Atomic Approaches to Defect Thermochemistry

5. FUNDING NUMBERS
G - AFOSR-89-0309

6. AUTHOR(S)
Professor James A. Van Vechten
Associate Professor John F. Wager

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Oregon State University
Dept. Electrical & Computer Engineering
Corvallis, OR 97331-3211

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 92 0782

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR /NE
BOLLING AIR FORCE BASE
DC 20332

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
2306B1

DTIC ELECTE AUG 17 1992 S

11. SUPPLEMENTARY NOTES

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

12a. DISTRIBUTION/AVAILABILITY STATEMENT
unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We have achieved insight into the role of H in semiconductor crystal growth in processes such as organometallic chemical vapor deposition, OMCVD. If allowed, H will compensate shallow dopants and suppress the formation of other compensators when the Fermi level moves significantly from the intrinsic level at growth or processing temperatures, T. This can be a great advantage because the H can be removed at low T with no rearrangement of other atoms. This has led to the first attainment of good p-type GaN and to improvement of n-type GaN. It should work for any semiconductor. We have improved the thermodynamic analysis of heterojunction band offsets as functions of T and strain. We have done the first fundamental studies of the diffusion equation with boundary conditions appropriate for crystal growth and diffusion and obtained major insights. We devised a simple new experiment to study interstitial impurity diffusion in semiconductors with striking results. We demonstrated that host interstitials play no role in thermal in Si or GaAs and similar low ionicity crystals and have explained the "U-shaped" profile of transition metal impurities. We have verified that the DX center can be getter out of AlGaAs without reducing donor concentration and studied its properties by various capacitance transient. We have characterized the aging and charge trapping properties of ZnS ACTFEL display materials. We have calculated the entropy of atomic hopping.

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

FINAL TECHNICAL REPORT

to

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE
WASHINGTON, DC 20332

on

ATOMIC APPROACHES TO DEFECT THERMOCHEMISTRY

(AFOSR-89-0309)

for period

1 April 1989 to 31 March 1992

Submitted by

Professor James A. Van Vechten

and

Associate Professor John F. Wager

Department of Electrical & Computer Engineering
Center for Advanced Materials Research
Oregon State University
Corvallis, OR 97331-3211

ABSTRACT

The goal of the research program described herein was to apply atomistic thermodynamic theory, Monte Carlo simulation, and experimental analysis to elucidate the identity of point defects as well as to understand their static and dynamic properties. Significant progress has been made in the following eight areas:

I. H as a Removable Compensator in GaN and Other Semiconductors

II. Thermodynamic Analysis of Semiconductor Band Offsets

III. Monte Carlo Simulations of Atomic Diffusion and Interactions

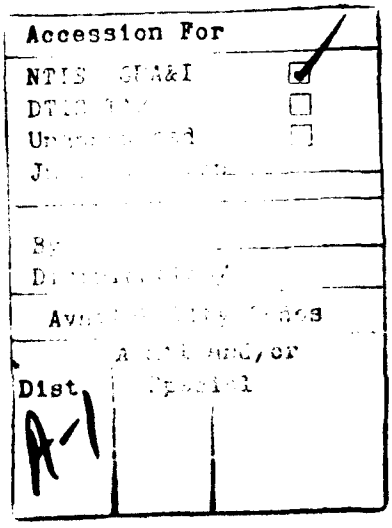
IV. Controlled Experiments to Study Transition Metal Diffusion and to Map Vacancy Concentrations at a Fixed Time

V. Studies of Electroluminescent Flat-Panel Display Devices

VI. Defect Characterization

VII. Atomistic Thermodynamic Theory and Applications

A summary of this work is as follows, while a detailed discussion is available in the publications enclosed with this report.



I. H as a Removable Compensator in GaN and Other Semiconductors

1. **J.A. Van Vechten, "A Simple Man's View of the Passivation of Semiconductors," *Solid State Electronics 33*, 39 (1990).**
2. **J.A. Van Vechten, J.D. Zook, R.D. Horning, and B. Goldenberg, "Defeating Compensation in Wide Gap Semiconductors by Growing in H that is Removed by Low Temperature De-Ionizing Radiation, " *Jpn. J. Appl. Phys.* in press (1992).**

The perhaps most relevant insight from the investigations made under this grant occurred towards its end. It was done in collaboration with Barbara Goldenberg, Robert D. Horning and J. David Zook of Honeywell Sensors and Systems Development Center, Bloomington, Minnesota. It was presented at the Wide Bandgap Nitrides Workshop held by the Office of Naval Research (Max Yoder) and the University of Illinois (Hadis Morkoc) in St. Louis, Missouri, on 13 April 1992. A six page summary of that presentation was distributed at that workshop. The workshop was attended by Prof. I. Akasaki of Tskuba University (Japan) and Dr. S. Nakamura of Nichia Chemical Industries, Tokushima, Japan, who had published the first reports of successful production of highly p-type GaN, which called our attention to the question how it may be possible to defeat the tendency of widegap semiconductors to self compensate. It is clear that the groups of Akasaki and of Nakamura have done a remarkable job of this in the case of GaN, which has been attempted by many groups around the world for more than 25 years. Neither Akasaki nor Nakamura have published an explanation why their method, which has not been fully disclosed, should have succeeded where so many others have failed. JAVV presented a general prescription how the tendency of crystals to compensate dopants during processing could be defeated by deliberately close compensating with H ions (H+ to make p-type or H- to make n-type) during growth at low temperatures (e.g., by OMCVD), so that the system would not have the normal thermodynamic driving force to compensate shallow acceptors or shallow donors with native defects or with unwanted impurities, and then removing the H ions by low temperature, low energy processes when high temperature processing has been completed. Thus, the shallow dopants are incorporated into unusually high quality crystal (because many fewer antisite defects, interstitials and vacancies were grown in due

to the removal of compensation as a motivation for their formation) and are uncompensated once the H ions have been removed. This method also increases the solubility of the shallow dopants in the host crystal. We asserted that this is a very generally method that can improve conductivity of either type in any semiconductor and that Akasaki and Nakamura had produced their p+ GaN by practicing our general prescription for that specific case.

The St. Louis presentation provoked a strong response especially from Dr. Nakamura, who disclosed that he has submitted more information and explanation for publication in the *Japanese Journal of Applied Physics*. It seems that this information is in complete accord with our theory and prescription and that Dr. Nakamura had independently formed a similar theory, but had not realized how broad the implications are. It was also evident that few of the other attendees of the workshop had considered the role of the H which is generally present in state-of-the-art crystal growth processes. Our collaborators at Honeywell SSDC grew two samples of GaN, one doped with donors and the other with acceptors, by OMCVD using NH3, which cracks on the sample surface to supply atomic H. As grown, the samples were weakly n-type and semi-insulating respectively. A simple heat treatment was then used to drive out the incorporated H- or H+ ions. This converted the n-sample to n+ and the semi-insulating sample to p-type, 3 × E17 with a mobility of 7, which is within an order of magnitude of the best material reported by the Japanese. We feel this recent result confirms our theory for the removable H compensator method.

After considerable discussion of the issues of patents and trade secretes with the Honeywell group and their management, we decided to take purely defensive but strong position. Thus, we presented this insight at the St. Louis workshop and submitted the 6 page summary for publication as a "Short Note" in the *Japanese Journal of Applied Physics*, where Dr. Nakamura's manuscripts, and perhaps others, are being reviewed and prepared for publication. We feel that by acknowledging receipt of our manuscript *JJAP* acknowledges that the broad concept is now published in Japan. From now on attempts to patent processes such as Akasaki and Nakamura have developed can be attacked as obvious in view of our

publication. Honeywell and other U.S. manufactures should have freedom of action to use the general method without paying royalties.

At this writing, we are preparing to pursue the subject further at the Materials Research Society Meeting in San Francisco (27 April 1992) where several Japanese repoits on the role of H in high conductivity semiconductors are to be presented. We will also prepare a manuscript with a detailed and quantitative discussion of the issue for publication in a U.S. journal. A copy of the St. Louis summary is attached.

II. Thermodynamic Analysis of Semiconductor Band Offsets

1. **J.A. Van Vechten and K. J. Malloy, "The Temperature Dependence of Band Offsets for Semiconductor Heterojunctions in General and for the Particular Cases of AlAs-GaAs and HgTe-CdTe," *J. Phys.: Condens. Matter 2*, 281 (1990).**
2. **J.A. Van Vechten, "Atomic Diffusion with Strain and Injection," *Mat. Res. Soc. Symp. Proc. 184*, 165 (1990).**
3. **K.J. Malloy and J.A. Van Vechten, "Thermal Expansion Contributions to Band-Gap and Band Offset Temperature Dependences," *J. Vac. Sci. Technol. B 9*, 2212 (1991).**

The dramatic advance in semiconductor heterojunction technology has created a motivation to understand the magnitudes of the valence band offset and the conduction band offset at any particular heterojunction and their variations with temperature and with strain. There have been several rather involved electronic structure calculation treatments of these questions and these leave much controversy in the interpretation of a very large body of experimental reports. The controversy was particularly severe for the cases of AlAs/GaAs and CdTe/HgTe.

We have applied a thermodynamic approach to the band offset problem which makes use of rigorous thermodynamic definitions of the electron affinity, work function, and valence band ionization potential. When discussed in terms of these thermodynamic potentials, it is clear that conduction band minimum and the valence band maximum are bulk thermodynamic electro-chemical potentials. Thus, the quantities defined in this way must be orientation dependent and transitive from one junction to another. This permits a great simplification of the general problem and particularly of the questions of

temperature and pressure variation. However, care must be taken in the analysis of experimental data because many works do not treat their data as thermodynamic data. The use of this approach is often called the use of the "Anderson model"; it is controversial because not everyone agrees to use the thermodynamic approach, nor understands how.

Kevin J. Malloy, now at U. New Mexico, and JAVV have shown how to use this approach to resolve the experimental controversy re the band off-sets in AlAs/GaAs and in CdTe/HgTe in particular, as well as to treat the general problem. They have also treated the strain problem and the contributions of thermal expansion to band-gap and band-off sets. The latter is relevant to the development of, e.g., the self-electrooptic-effect detector, SEED, for implementation of "neural network" devices, where variations of optical emission and absorption peaks with unpredictable temperature fluctuation induced the operation of the device is a major problem. One may hope to resolve this problem by engineering a package with materials having different coefficients of thermal expansion so as to produce stresses that buck-out the pure thermal variations. It is also relevant to the issue diffusion of charged species near a heterojunction.

III. Monte Carlo Simulations of Atomic Diffusion and Interactions

Using the VIDSIM vacancy and interstitial diffusion simulator developed early in this grant, we have done direct, atom level, Monte Carlo simulations of four diffusion problems with realistic boundary conditions and various initial conditions. These are: i) the first and second neighbor hopping diffusion of vacancies in GaAs and AlGaAs heterostructures and alloys; ii) the "kick-out" mechanism for a interstitial impurity, especially a transition metal, diffusing into an originally perfect semiconductor (e.g., Si or GaAs); iii) the "Frank-Turnbull" mechanism for the same process as ii) above; and iv) diffusion of a native defect, e.g., a vacancy, from a surface with an annihilating boundary condition, e.g., a free surface, into an initially perfect crystal with a marker layer of either host or impurity atoms at various depths. Several tens of billions of atomic events in crystals typically containing 6.4E8 lattice sites have

now been simulated by running VIDSIM on up to 28 microcomputers when these machines would otherwise be idle. (Thus, at no cost.)

i) found that the rate of Al diffusion across an AlAs/GaAs heterostructure into GaAs should depend not only on doping level but also on the choice of dopant. This is because both Al and the dopant atoms interact with the hopping vacancies so that the hops of these atoms are not uncorrelated events. When the dopant atom is light, easy to move and easily placed on either sublattice, e.g., Si in GaAs, then it tends to migrate with the Al and the two attract a cluster of vacancies. When the dopant is heavy, hard to move and much more willing to sit on one sublattice than the other, e.g., Te in GaAs, Then the Al tends to leave the Te behind and many fewer vacancy clusters are formed. We deduced that Te doping should produce an Al diffusivity proportional to n**1 exp(-2.9 eV/kT) while Si doping should produce a diffusivity proportional to n**3 exp(-4 eV/kT), where n = ND - NA is the doping density, about the same time that P. Mei et al. published the first experimental report of this effect (*Appl. Phys. Lett. 53*, 2650 (1988)).

ii) For the last dozen years or so, several groups around the world have sought to explain the distribution profiles of substitutional transition and noble metal impurities obtained in semiconductors by diffusion. These profiles differ from those obtained for shallow dopant impurities, which further to the right in the Periodic Table. The transition and noble metals spread much deeper much more rapidly and are often found to produce an almost symmetric, two sided profile, which is called "U shaped" or "Bath Tub shaped", from a one sided source. The case that has been most discussed in the literature is that of Au diffusing into Si, but Cu, Ni, Pt, and Pd diffusion into Si and all these diffusing into Ge and GaAs have also been reported and remarked upon. In the case of GaAs, the two sided profile from a one sided source feature has not been reported but the observed near side profile is more like that for Si or Ge than is the profile for a shallow dopant impurity diffusing into GaAs.

The original interpretation of these profiles is that diffusion is occurring with the "Frank-Turnbull mechanism", abbreviated FT. It was assumed that the transition or noble metal, M, diffuses as an

interstitial but occupies substitutional sites when it encounters a vacancy, V. It was suggested that the bath tub shaped profile resulted from a one-sided source because V's diffused slowly from both sides of the sample, which originally was perfect, while M diffused so rapidly from one side a bulk equilibrium value for the M interstitial, M_i, was soon established in the cases of Si and Ge. For GaAs, it was assumed that there were so many sinks for the M_i in the bulk of the sample (dislocations or perhaps V's) that the M_i did not get to the far side.

The FT model was challenged by Dr. A. Seeger (MPI Stuttgart) and his students, particularly U. Goesele, who claim it cannot account fit the time variation of the substitutional concentration, [M], in the flat central region of the bath tub profile nor fit the details of the shoulders on the two sides. They asserted that instead a "kick out mechanism", KO, must be dominant. They asserted that M_i's diffuse rapidly into initially perfect crystal and, at least for the case of Si, soon establishes a uniform concentration at the bulk equilibrium value. They postulated that the M_i's "kick" the host atoms out of their substitutional sites to produce the substitutional M impurities and host interstitials. They neglected any effect of vacancies and assumed that the host interstitials either annihilate at the free surface of the sample or at dislocations if any are present or they recapture the lattice site from the M substitutionals. Thus, they assert that the rise of [M] is governed by the out diffusion of the host interstitials rather than by the in diffusion of vacancies, as assumed in the FT model; both groups assumed that the sample was initially perfect, i.e., had no vacancies and no self interstitials.

Our direct Monte Carlo simulation of both the FT and the KO models, as originally proposed, found that in fact neither of them resemble the reported data that they were supposed to fit. We have been able to establish why the discrepancies exist.

In the case of the KO model, there is a fundamental inconsistency between the assumption that the M_i concentration, $[M_i]$, is uniform throughout the sample very shortly after the onset of the diffusion and assumption that the M_i's can displace host atoms at a significant rate. We found, and in retrospect it seems quite obvious, that if M_i can displace host atoms and if M_i's diffuse into a region where initially

[M] = 0, then [M_i] will decay into that region. The exact behavior of [M_i] and [M] with time depend upon the boundary condition assumed at the far side of the sample and on the ratio of the mean diffusion path of the M_i before it displaces a host atom. We tried the simulation with both the assumption that the M_i are annihilated at the far surface (as would occur if they evaporated from there) and the assumption that they are reflected back into the sample from the far surface (as might occur if an alloy of M with the host forms are the far side). We also tried a range of from 1.0

to 20 for the ratio of the sample thickness to the mean diffusion distance before a displacement. In all cases for the one sided source, both [M] and [M_i] decreased monotonically from the source side to the far side for the duration of net diffusion, i.e., until [M] saturated. Saturation occurred first at the source side and spread monotonically across the sample. See Schmid et al. *Mater. Res. Soc. Symp. Proc. 163*, 609 (1990) and Van Vechten et al. *J. Electron. Mater. 20*, 431 (1991).

We also found that for two sided sources, the KO model with an initially perfect sample does not give the [M] profile claimed by it advocates. Again the reason is that, contrary to their assumption the [M_i] profile is not flat from very early times onward.

iii) We also did extensive simulation of the FT model under the assumptions that the sample is initially perfect and that any vacancy or M_i that crosses out of the sample through either free surface is annihilated. The latter corresponds to the widely held opinion that the free surface is "an ideal source and sink" for all native point defects. (We will return to this ill founded opinion later.) For the case of a one side source, contrary to the assertions made by its advocates, our simulation shows that this model does not produce a symmetric U shaped profile. It does produce a shoulder on the source side, a rather flat minimum through the center, and a broad plateau on the far side, but with a dip toward the far surface. Again, the reason for this result is obvious in retrospective. Although the

concentration of vacancies, [V], has a shoulder with a maximum at the far surface, similar to that at the near surface, the M_i live only a short time in the near surface region before they annihilate. Thus, the M_i's have little chance to find a V near the far surface under the assumption that the source is one

sided. Deeper into the far surface, [V] is lower but the M_i's live much longer so they are more likely to form a substitutional. The balance between the decreasing [V] and increasing life expectancy for the M_i with increasing depth from the far side produces a plateau between the central region, where [V] is almost zero so there are very few M substitutionals, and the near-far-surface layer that has no M substitutionals.

We also found that this FT model with the assumption that the sample is initially perfect does not produce the "square root of time" variation for [M] in the middle of the sample that several groups have reported.

Thus, we demonstrated with our direct Monte Carlo simulation that neither of the two commonly advocated mechanisms for transition metal diffusion into semiconductors (the KO and the FT with the assumptions described above) can in fact account for the data commonly reported in the literature. As described in the next section, we have established with experiments on Si that the resolution of the question of the mechanism of transition metal diffusion into Si flows from the realization of the facts that:

1) The surfaces of typical samples are not ideal sources and sinks for native point defects when undisturbed;

2) The actual rates of diffusion of these metals into the semiconductors is very sensitive to the actual condition of the surfaces of the sample, i.e., to the density of kink sites on the surfaces;

3) The "U shaped" data typically reported in the prior literature is obtained only when there is such a large excess of metal on the source side surface that an alloy layer forms also on the far side surface;

4) Even for Si, the samples are not initially perfect. Indeed, they contain the residue of the vacancies that were grown into the boule that the 1685 K growth temperature because they have no chance to diffuse out of any but the near surface region during wafer processing. According to our best estimate, and also that of researchers at

Philips and SEH, about 8E17 vacancies per cc are trapped into about 2E9 voids per cc of typical diameter 200 nm, which are called D defects and are easily observed after decoration with Cu, Li, Na, etc. by x-ray tomography, or without decoration but with much difficulty by electron microscopy.

5) The shape of the profile of substitutional M upon diffusion into Si or other semiconductors is qualitatively different if the M-host alloy melts on the source side surface, and also where M has decorated the D defects, than if these alloys are solid. The temperature at which these alloys melt varies with the choice of metal (and host). What has been claimed to be a transition from diffusion dominated by the FT mechanism at low T to diffusion dominated by the KO mechanism at high T is simply the transition from having the alloy solid to having it molten.

6) When account is taken of the effect of the voids initially present the sample, the FT mechanism predicts the same "U shaped" profile as has been ascribed to the KO model under the conditions that this profile is in fact obtained. See R.K. Graupner et al. *PCSI-19* and *MRS Proc.* 4/92 in press.

iv) When we assume that the free surface of the sample is an ideal source and sink for native defects, i.e., that they are created there at random sites at random intervals with an average rate that depends only upon temperature and that they are annihilated if they diffuse back out through the surface, we find that, whatever other assumptions are made, our simulation gives us a dip to zero for the concentrations of any of the native defects at any free surface. This should have been obvious as a consequence of the facts that these concentrations must be zero one step beyond the surface and that the diffusion equation implies continuity in the concentrations of the diffusing species. However, the literature of the field is full of paper that assert as an initial condition that the defect concentrations at the free surfaces are equal to the bulk equilibrium value for the ambient temperature.

The source of this misconception in the literature is the classic treatments of diffusion for the cases of Brownian motion and of coagulation by Einstein and by Chandrasekhar. They used conservative boundary conditions, i.e., no diffusers were created and none were destroyed; if the diffuser stepped out through the surface of the sample under consideration, it has a high probability to step back in the near future. In the case of transition metals diffusing into a semiconductor, if the defect passes out the free surface, then it is gone and another will be created only after the random creation interval and then it will appear at a random site not likely to be close to where the previous defect was annihilated. Einstein and Chandrasekhar derived the error function complement formulae that are quoted in many text books. It is unfortunate that so many workers have assumed these formulae for semiconductor diffusion problems without examining their applicability. We all know that if you change the boundary conditions, you change the solution to a differential equation.

We have determined that there is no closed form, analytic expression to describe the diffusion profile of a point defect from a free surface with annihilating boundary conditions into an initially perfect sample. Therefore, we have used our VIDSIM simulation program to generate directly the statistics that describe such a profile and compare and contrast them to the Einstein-Chandrasekhar, EC, result for the conservative boundary condition. We have generated about 15 billion events worth of statistics for this fundamental study. We have found some results which we regard as striking. These include: i) the mean square penetration depth increases linearly with time, as in the EC case, but the EC value Dt (where D is the diffusivity and t the time of diffusion) is multiplied by a constant factor that is close to 2/3; ii) far from the surface (where there is the dip mentioned above) the profile can be fitted very well to the error function complement formula that EC obtained if Dt is again multiplied by a constant factor that does not seem to vary with t. See Zhang et. al.

IV. Controlled Experiments to Study Transition Metal Diffusion and to Map Vacancy Concentrations at a Fixed Time

1. J.A. Van Vechten and U. Schmid, "Vacancy First and Second Neighbor Hopping at a Compound Semiconductor Interface: Insights from Computer Simulation," *J. Vac. Sci. Technol. B 7*, 827 (1989).
2. U. Schmid, J.A. Van Vechten, N.C. Myers, and U. Koch, "Failure of the "Kick-Out" Model for the Diffusion of Au into Si when tested by Monte Carlo Simulation," *Mater. Res. Soc. Symp. Proc. 163*, 609 (1990).
3. J.A. Van Vechten, U. Schmid, and Zhang Q.-S., "Surface Treatment Effects on Atomic Diffusion in Si Explained without Self-Interstitials," *J. Electron. Mater. 20*, 431 (1991).
4. R.K. Graupner, J.A. Van Vechten, and P. Harwood, "Characterization of Point Defect Generation at Silicon Surfaces using Gold Diffusion," *Proceedings of PCSI-19* 1/92 to be published in *J. Vac Sci. Technol. B* July/August 1992.
5. R.K. Graupner, J.A. Van Vechten, P. Harwood, and T.K. Monson, "The Effect of Vacancies Grown into Silicon on Gold Diffusion," *Proceeding of the Materials Research Society Symposium* 4/92, to be published 1992.
6. Zhang Q.-S., Zhao P., and J.A. Van Vechten, "Consequences of an Annihilating Boundary Condition on Point Defect Diffusion into an Initially Perfect Sample," APS March Meeting 92 and to be published.

We determined that, in order to understand how the commonly reported "U shaped" or "bath tub" profiles for transition metals diffusing from one side into semiconductors (particularly Si) could occur when neither the KO nor the FT model for diffusion into a initially perfect sample from one side produces this result, we needed to do some diffusion experiments of our own in collaboration with state-of-the-art producer of the relevant materials. (We formed a collaboration with Komatsu Si USA and Wacker Siltronic, both of Portland, Oregon. SEH America of Vancouver, Washington, has now joined this collaboration.) At the same time we realized that the fact that the transition metals diffuse so rapidly and that they primarily occupy substitutional lattice sites, could allow us to map out the vacancy concentrations in the sample at an arbitrary fixed time in an annealing sequence.

The previous diffusion experiments reported in the literature all involved depositing a layer of the metal on the surface prior to the diffusion annealing. They also all performed the diffusion either in a diffusion furnace with flowing gases or in a (partially) evacuated amphule; in either case there is at least 1 torr of oxygen, O, present.

We arranged to do the diffusion in an ultra-high vacuum chamber where we could either have hard vacuum or controlled amounts of O or other gases present. We further provided that the metal to be diffused would be deposited in precisely measure quantities onto the sample at any time before or after it had reached diffusion temperature. Because the M_i diffuse so rapidly, we found that 5 minutes is sufficient time at 960 C to diffuse the M_i's completely through a 0.5 mm thick Si sample. It would take more than 100 hours to equilibrate the vacancies at this temperature. Thus, if we anneal the sample any way we choose, we can then deposit the metal and hold the temperature for 5 minutes, then cool the sample to room temperature radiatively in another minute and know that a large fraction of the single vacancies that were present when the M arrived are now occupied by M substitutional impurities. The substitutional M profile is easily measured by the spreading resistance profiling, SRP, method.

As already noted above, some startling results came from these studies. These include: i) the profile is qualitatively affected by the state, molten or solid, of the alloy formed by the metal on the host; ii) the far side profile is dramatically affected by the presence of as much as 0.1 torr of O; iii) much of the M that enters the sample does not appear as electrically active M substitutionals; instead, it decorates the D defects, i.e., it forms (molten or solid) alloy specs at these voids; iv) One does not get the oft reported bath tub shaped profile unless one deposits enough M on the source side that enough M_i get to the far side to form an equivalent alloy source there. The reason that the profiles are so perfectly symmetric is thus obvious; contrary to what was claimed, they in fact are produced from a two sided diffusion. Furthermore, the alloy must be molten to get the "U shaped" or bath tub profile.

V. Studies of Electroluminescent Flat-Panel Display Devices

1. **I. Khormaei, "Improved Stability of ZnS:Mn ACTFEL Devices," M.S. Thesis, Oregon State University (1989).**
2. **I. Khormaei, J.F. Wager, and C.N. King, "Improved Stability of ZnS:Mn ACTFEL Devices," *SID 89 Digest*, 65 (1989).**
3. **R.C. McArthur, J.D. Davidson, J.F. Wager, I. Khormaei, and C.N. King, "Capacitance-Voltage Characteristics of AC Thin Film Electroluminescent Devices," *Appl. Phys. Lett. 56*, 1889 (1990).**

4. R.C. McArthur, J.D. Davidson, J.F. Wager, I. Khormaei, and C.N. King, "Characterization of ZnS:Mn AC Thin-Film Electroluminescent Devices by Capacitance-Voltage Analysis," *Acta Polytechnica Scandinavia Ph 170*, 181 (1990).
5. J.D. Davidson, J.F. Wager, I. Khormaei, and C.N. King, "Aging Instabilities of ZnS:Mn AC Thin-Film Electroluminescent Devices," *Acta Polytechnica Scandinavica Ph 170*, 185 (1990).
6. J.D. Davidson, I. Khormaei, and J.F. Wager, "Electrical Characterization and SPICE Modeling of ZnS:Mn ACTFEL Devices," *SID91 Digest*, 77 (1991).
7. I. Khormaei, C.N. King, R.E. Coovert, and J.F. Wager, "Stabilization of ZnS:Mn ACTFEL Devices Through Processing Modifications," *SID91 Digest*, 74 (1991).
8. J.D. Davidson, "Capacitance-Voltage Analysis, SPICE Modeling, and Aging Studies of AC Thin-Film Electroluminescent Devices," M.S. Thesis, Oregon State University (1991).
9. J.D. Davidson, J.F. Wager, I. Khormaei, C.N. King, and R. Williams, "Electrical Characterization and Modeling of Alternating-Current Thin-Film Electroluminescent Devices," *IEEE Trans. Electron Devices ED-39*, 1122 (1992).
10. J.D. Davidson, J.F. Wager, and S. Kobayashi, "Aging Studies of ZnS:Mn Alternating-Current Thin-Film Electroluminescent Devices," *J. Appl. Phys. 71*, 4040 (1992).
11. J.F. Wager, A.A. Douglas, and D.C. Morton, "Electrical Characterization and Modelling of ACTFEL Devices," *EL-92, Sixth International Workshop on Electroluminescence*, El Paso, TX, May 11-13, 1992.

Alternating-current thin-film electroluminescent (ACTFEL) devices are emerging as a viable flat-panel display technology with potential for applications such as high-definition television. Our ACTFEL work has focussed on electrical characterization and modeling, strategies for improvement of the device stability, and aging studies.

With respect to electrical characterization and modeling, we proposed and have developed and refined the capacitance-voltage (C-V) technique for the electrical characterization of ACTFEL devices. We find this technique to be complementary to the charge-voltage (Q-V) which is conventionally utilized for ACTFEL characterization. SPICE modeling of ACTFEL devices has also been employed in conjunction with C-V and Q-V analysis. We have found C-V and Q-V analysis in conjunction with SPICE modeling to be a powerful approach for understanding the device physics and aging characteristics of ACTFELs. Recently we have refined the SPICE model to account for parasitic resistances within the ACTFEL device and to more closely account for details of the ACTFEL dynamic response. Additionally, we have refined the Q-V technique by defining two new quantities, leakage charge, Q_{leak}, and relaxation charge, Q_{relax}; these new quantities have enabled us to understand trends in the ACTFEL response as

a function of changes in the excitation voltage and have allowed us to begin to attack the problem of understanding the device physics underlying ACTFELs with asymmetrical interfaces.

We have investigated three successful approaches for improving the brightness-voltage (B-V) stability of ACTFEL devices. One approach involves the addition of a CaS layer at one or both of the phosphor/insulator interfaces. Another approach is to utilize oxygen exposure of ZnS prior to the second insulator deposition. A third approach is to employ post-deposition sulfur anneals. We believe that B-V instabilities are associated with sulfur vacancies.

Aging studies of evaporated ZnS:Mn devices were recently reported. Q-V and C-V analysis lead to a picture for ACTFEL aging in which atomic arrangement at the insulator/phosphor interfaces, given rise to the formation of deep level, fixed charge states. These fixed charge states capture conduction electrons and perturb the internal electrostatics in such a manner to account for the observed aging instabilities.

VI. Defect Characterization

1. **T.W. Dobson, "Transient Decay of Persistent Photoconductance in AlGaAs," M.S. Thesis, Oregon State University (1989).**
2. **T.W. Dobson, L.V.A. Scalvi, and J.F. Wager, "Transient Decay of Persistent Photoconductivity in $Al_{0.3}Ga_{0.7}As$," *J. Appl. Phys. 68*, 601 (1990).**
3. **S.B. Kim, "Electrical Characterization of n-type Aluminum Gallium Arsenide," Ph.D. Thesis, Oregon State University (1991).**
4. **S.B. Kim and J.F. Wager, "Low-Temperature Hole Capture Cross Section of the DX Center in AlGaAs," *Semi. Sci. Technol.* (submitted).**
5. **C. Huang, "A Study of Deep Levels in AlGaAs/GaAs Heterojunction Bipolar Transistors," M.S. Thesis, Oregon State University (summer 1992).**

Two experimental investigations were undertaken with the goal of clarifying the atomic nature of the DX center in AlGaAs. The first experiment is denoted transient decay of persistent photoconductivity (TDPPC) and consists of monitoring the PPC transient as a function of temperature and simulating the experimental data assuming thermally activated electron capture into DX with a concomitant modification

of the ionized impurity density. Six models for DX were considered but the negative-U model for DX of Chadi and Chang in conjunction with a shallow donor led to the most satisfying results.

A second investigation involved the minority carrier capture (MCC) technique to estimate the low-temperature capture cross section of minority holes into the DX center. The measured magnitude of the low-temperature capture cross section is interpreted to be consistent with the negative-U model for DX of Chadi and Chang if the rate-limiting step for hole capture involves DX in a neutral charge state.

Additional work, currently in the final stages of completion, involves the electrical characterization of AlGaAs/GaAs heterojunction bipolar transistors.

VII. Atomic Thermodynamic Theory and Applications

1. **T.W. Dobson, J.F. Wager, and J.A. Van Vechten, "Entropy of Migration for Atomic Hopping," *Phys. Rev. B40*, 2962 (1989).**
2. **T.W. Dobson and J.F. Wager, "Enthalpy of Formation of Antisite Defects and Antistructure Pairs in III-V Compound Semiconductors," *J. Appl. Phys. 66*, 1997 (1989).**
3. **J.F. Wager, "A Statistical Thermodynamic Derivation of the Ballistic Model for Vacancy Migration," *Phil. Mag. A 63*, 1315 (1991).**
4. **J.F. Wager, "Energetics of Self-Diffusion in GaAs," *J. Appl. Phys. 69*, 3022 (1991).**
5. **J.F. Wager, "Thermodynamics and Kinetics of Vacancy Self-Compensation in Wide Band Gap Semiconductors," *Phil. Mag. A* (accepted).**

Atomistic thermodynamics is the application of macroscopic thermodynamic principles to the study of materials on the atomic scale. We have both refined and applied atomistic thermodynamic theory in the work cited above.

Two main refinements of atomistic thermodynamic theory have been achieved. First, a formulation of the entropy of migration of an atom undergoing a nearest-neighbor vacancy hop was presented and was derived by statistical thermodynamics. The second refinement was a reformulation of the enthalpy of formation of neutral, isolated antisite defects and antistructure pairs in compound semiconductors.

Atomistic thermodynamic theory was applied to a number of semiconductor materials problems. First, the entropy of migration formulation was employed in order to explain InP drain-current-drift measurements, deep-level-defect-transformation kinetic studies of the metastable M center in InP, in Si

and Ge self-diffusion experiments,and in elemental metal self-diffusion experiments. Energetic and entropy considerations were used to examine atomic mechanisms of self-diffusion in GaAs. Finally, an atomistic thermodynamic formulation was employed in order to explain self-compensation trends in wide band gap compound semiconductors.

Corrosion Science. Vol. 31, pp. 39–52, 1990
Printed in Great Britain

0010–938X/90 $3.00+0.00

A SIMPLE MAN'S VIEW OF THE PASSIVATION OF SEMICONDUCTORS

J. A. Van Vechten

Center for Advanced Materials Research
Department of Electrical and Computer Engineering
Oregon State University, Corvallis, OR 97331-3211 U.S.A.

ABSTRACT

The author here attempts to present a fundamental and general discussion of the problem of passivating an electronic device made of semiconducting materials. He also points out what he believes to be a few misconceptions in the current literature and makes a suggestion for further progress.

KEYWORDS

Passivation, Semiconductor, Gettering, Oxide, Nitride, SiO_2, Hydrogen, Noble Gas, Dangling Bond

I. INTRODUCTION

We define the passivation of a semiconductor as the elimination, or at least the severe reduction, of chemical and electrical reactivity either at a particular interface, such as its surface, or through its bulk.

Manufacturers of electronic devices must achieve and maintain proper control of passivation if they are to attain any significant yield, performance and reliability for their products. This being the case, a great many workers have explored many strategies in hopes of achieving this goal. A diverse literature has arisen. The author will here try to provide a simple and coherent general formulation of the problem and review the current state of the subject.

It has long been widely agreed that the electrical passivation of common, practical semiconductors (i.e., Si, GaAs, GaP, InAsP, etc.) is equivalent to their chemical passivation. This implies that electrons and holes in their band edge states do not self-trap via lattice distortions in these materials, as they often do in alkali halide crystals, and do not themselves cause chemical reactions, as they do in Ag halides to produce the photographic process. On the basis of certain calculations made with what is known as the local density functional approximation, LDA, several theoretical physicists have recently challenged this premise (1,2). In particular, they claim that the low temperature passivation of donor impurities in AlGaAs alloys, which is known as the DX phenomena, is not a consequence of a chemical reaction involving some lattice defect, X, but a pathological property of all substitutional donors in these alloys. If this were true, it would be a very discouraging result for the whole technology because it would imply there is no hope of producing devices with n-type conductivity in these alloys at moderately low temperatures, T. Many economic analyses

conclude that AlGaAs devices can compete successfully against Si devices with similar function only if they can be made to operate below T = 135 K, where Si bipolar transistors cease to amplify. Thus, the conclusion implied by the LDA theory calculations is that most research and development of the AlGaAs alloys should be terminated. LDA calculations ascribe the "EL2" phenomenon in GaAs and a few other important passivation issues to a similar pathological property of isolated atoms. The author is convinced that in this case the traditional premise is correct and the recent LDA theory claims are quite wrong. In the Section II he gives theoretical and empirical arguments to support his conviction.

If the reader will accept the traditional premise that the passivation of the practical semiconductors involves the properties of the "dangling" or reactive bonds of these covalently bonded solids, which are present around certain impurity atoms and lattice defects, rather than the inherent properties of good bonds among host atoms and dopant atoms, then control of passivation means control of dangling bonds.

In principle it seems there are two aspects of passivation - mechanical and electronic. However, in practice they seem to be linked; where a semiconductor is mechanically metastable, i.e., where atoms can move, there generally are localized electronic levels in the gap; the converse is also true. Let us enumerate possible strategies. i) One can provide a barrier layer, e.g., SiO_2, to passivate the semiconductor in the same way that Al_2O_3 passivates metallic Al. In Section III we consider the requirements of a good barrier layer and why SiO_2 is so satisfactory for Si and rather less adequate for III-V semiconductors. ii) One can terminate the dangling bonds with H. Many workers have reported interesting problems and possibilities in this direction recently. The subject is briefly commented upon in Section IV. iii) One can getter objectionable dangling bonds out of regions where they cause trouble to regions where they present less problem or may even provide a benefit. Gettering generally implies use of strain fields. We discuss gettering in Section V. iv) One can use elements other than H to terminate dangling bonds, or otherwise depresses undesired reactions. This possibility is discussed in Section VI and the use of noble gas atoms for the purpose is suggested. Some concluding remarks are made in Section VII.

II. PASSIVATION AT DANGLING OR AT COMPLETED BONDS

The traditional premise that passivation is a process happening at dangling bonds has been challenged by calculations (1,2) using the LDA approach and the interpretation of some experiments (3,4). Although the challenge is more general, there has been a concentration on GaAs and AlGaAs alloys, and in particular on the EL2 and DX phenomena in these two. These phenomena are reactions between metastable states of these semiconductors of great industrial importance because they switch the material between a highly conducting state and an insulating state. Also, both reactions can be passivated (5,6) with H. (This passivation seems to be not practical for DX and only marginally practical for EL2.) They must be included in any broad discussion semiconductor passivation. Thus, a challenge to one's view of these reactions is a serious challenge to one's view of the entire subject of passivation. The author's response to this challenge is divided into three parts: principle, theory, and experiment.

As a matter of princple, the author opposes the suggestion that dangling bonds are not crucial to passivation in semiconductors because it contradicts a central lesson of materials science taught by the experience of the

semiconductor industry (7-9). After 30 fruitless years trying to make bipolar and field effect transistors from Cu oxides, PbS, and the like, progress was finally made when the pioneers gave up the notion that these materials were easy to work with and to integrate with metal conductor arrays and instead turned their attention first to Ge and then finally to Si.

Why do Cu oxides fail so completely where Si succeeds so well? The directional, covalent bonds in the Cu oxides are too weak and atoms are too ionic to prevent the lattice from responding to the electrostatic field produced by conduction band electrons or valence band holes that are trying to pass trough the sample. Not only is the mobility poor due to this lattice distortion, but the structure is too fragile to withstand the injection currents and bias fields required to operate a practical transistor. A prime material requirement for a practical semiconductor is that it must have strong, directional, covalent bonds so that the atoms will maintain their bonding network when large bias fields are applied and when large currents of electrons or holes are injected or accumulated. It also helps if the bonding has a low ionicity so that the interaction of the atoms and these fields and currents is relatively low. Si is a much better choice than PbS or Cu oxides because it has zero ionicity and assumes the very stiff diamond lattice. Although partly ionic and not as stiff as Si or Ge, GaAs is a much better choice for making a transistor than is ZnSe or CuBr.

Obviously, for a practical device structure, the dopant atoms, which normally complete all four bonds to their nearest neighbors, must maintain the lattice as do the host atoms. Thus, the suggestion that an isolated donor dopant atom in AlGaAs, in response to a high Fermi level at low T, would distort from its normal lattice site to produce the "DX centers" which compensate such samples, contradicts the basic lesson just noted and the fact that AlGaAs makes good transistors at room temperature.

The theoretical objection to the use of LDA calculations to support this suggestion is that the the LDA does not obtain reasonable and reliable values for the local effective fields, LEF. (The LEF is the instantaneous electric field acting on an ideal point charge; it varies widely from the macroscopic average field both spatially on the scale of the unit cell and temporally at frequencies corresponding to all the virtual transitions that contribute to the dielectric constant of the sample (10,11).) This problem with LDA calculations has been noted by others (12,13) previously. It is closely related to the problem that, although they do well with ground state properties such as crystal structure, lattice constant, phonons and bulk modulus (14), LDA methods do not calculate the band gaps of semiconductors correctly. (They calculate the band gap of Ge and GaAs to be zero (15).) It is a standard practice, although not always noted any longer, to correct the calculated band gaps to their empirical values when presenting the result of LDA calculations. This is usually done by a simple translation (called the "scissors operator"), although more elegant methods are also available. Unfortunately, the corrections required from the delocalized band states that define the band gap of a semiconductor are not the same as the LEF corrections appropriate for the tightly localized, deep level defect states relevant to passivation.

A good test for the accuracy of the LDA methods to calculate the LEF in a semiconductor is their estimate of the binding energy of a positron to a vacancy. The positron is among the best realizations of an ideal test charge; in a typical experiment (16,17) there is only one positron in the sample at a time so there is no orthogonalization problem. The most extensive LDA calculations (18) for positrons at vacancies in Si (and GaAs)

seem to be those of Puska et al. from the very reputable group in Stuttgart. The value they calculate for the lifetime of a positron in a Si vacancy is 254 to 261 ps (depending on charge state), which is not far from the empirical values, about 270 ps. This implies that they have the ground state properties of the vacancy - positron system reasonably accurately, as is usual for LDA methods. However, they calculate the binding energy, i.e., the energy to remove the positron form the center of the vacancy to the interstitial channel, which is the integral of the LEF along that path, to be just 1 eV. This would imply that in high temperature experiments observing positron trapping at vacancies to estimate thermal concentrations and properties of vacancies, there should be a substantial probability that the positron would detrap from a vacancy after being captured there before it annihilates. This would be detected in such experiments and has never been detected. From this empirical fact, one can deduce that the binding energy is actually ^{3} eV or greater (16,19).

Given that, for the rather ideal case of the binding energy of a positron to a vacancy in Si, a good LDA calculation estimates the energy required to move a unit charge from a vacant lattice site to an interstitial position with an error of 3 eV, it is hardly likely that similar methods can be relied upon to estimate the activation barriers for the reactions of the EL2 and DX phenomena, or other processes relevant to passivation, which are of the order of 0.2 eV. Indeed, good LDA calculations (2) estimate as only 0.2 eV the energy barrier to displace a normal donor in AlGaAs from its lattice site to an interstitial position, where it might become a multiple acceptor. There is a similar estimate (1) to displace an isolated As antisite defect in GaAs.However,in the author's opinion, these facts ought not to be accepted as compelling evidence against the traditional view that the AlGaAs lattice is quite stiff, so that compensation reactions and passivation processes occur only at dangling bonds.

Turning to experiment we might first note that differential thermal analysis, DTA, of electron irradiated GaAs (20) shows the net energy release when an As interstitial annihilates a vacancy to be 8.5 eV. The displacement activation energy must be greater than this. It is hard to believe the corresponding values for isolated donor atoms or antisite defects could be very much less. Thus, the GaAs lattice is indeed found to be stiff.

Positron annihilation experiments (17) also teach us that GaAs samples always contain concentrations of vacancies and vacancy complexes in excess of 10^{16} cm^{-3}. It must be remarked that these experiments only detect those positrons which are rapidly trapped into the vacancy cavity. Thus positively charged vacancies and vacancies bound in positively charged complexes are generally not detected in these experiments. The total vacancy concentration must be greater than that detected by the positrons. Complexes of vacancies with antisite defects account for the deviation from ideal stoichiometry of this and similar compounds (9,21), which often exceed 1 part per 1000 or several 10^{19} cm^{-3}. For example, if a GaAs sample is rich in As, complexes are formed such as $V_{Ga}As_{Ga}V_{Ga}$ (22,23), which has no net charge and is strongly bonded because V_{Ga} is a single acceptor and As_{Ga} is a double donor. Other configurations, including the EL2 family of complexes (24-26), also form. However, the facts that positron studies find so many vacancies while DTA studies find the energy released when host interstitials annihilate vacancies is so large convince this author that host interstitials are almost never found in these complexes.

Mossbauer experiments (27-30) also find that a large fraction of the Sn_{Ga} donors in GaAs have As vacancies as nearest neighbors. Thus, the compound semiconductors have large numbers of dangling bonds, in various configura-

tions, that may serve as the site of reactions that we are interested in passivating.

It must be noted that the experimental observation that the DX phenomenon is observed in non-alloyed GaAs under pressure, and not observed in the same sample without the pressure, has been argued (3) to be strong evidence that the phenomenon is a pathological property of the isolated donors rather than one of dangling bonds. As has been noted before, this would only be convincing evidence if it could also be shown that there where not enough vacancies, or other point defects, in these samples to account for the effect under pressure. (One agrees that new defects are not added as the pressure is applied.) In fact the positron and Mossbauer experiments indicate that this is not true; there are sufficient vacancies in the non-alloyed GaAs to account for the DX phenomenon with the reaction (29-31)

$$D^+ + V^0_{As} + 3\ e^-_c \rightleftharpoons D^+ + (V_M M_{As})^{3-} \ , \qquad (1)$$

where e^-_c denotes an electron in the conduction band distribution, M is a group III metal atom (Al or Ga), and D^+ is a donor atom which may or may not be a near neighbor to the vacancy. The reaction here is simply nearest neighbor hopping of an As vacancy, V_{As}, which is a deep donor and thus is neutral in n-type material.

Mossbauer and perturbed angular correlation, PAC, experiments offer an strong test for the proposal that reaction (1) accounts for the DX phenomenon (29,30). (1) implies that only one third as many V_{As} hop has e^-_c are removed. Thus, at least 2/3 of the donors in the sample should be unperturbed in their local environment by the DX compensation while 1/3 or less should be strongly perturbed by the presence of a strongly localized triple acceptor. This charge will produce a large isomer shift a the fraction of donors near the vacancy-antisite complex while the majority of donors will have the isomer shift characteristic of an isolated site in the perfect lattice. Such Mossbauer experiments have now been reported, both for the AlGaAs and for GaAs under pressure, and do show that a majority of donors remain unperturbed by the DX compensation reaction and a fraction about 1/3 move into the field of approximately three localized electrons when the sample is compensated.

Another experimental fact that should be mentioned in this regard is that it is possible to getter DX centers by causing a nearby dislocation to climb (32). This was observed in studies of the growth of "dark line defects" in AlGaAs double heterostructure lasers several years ago. As the active layer remains n-type, the donors remain while the X's, evidently V_{As}'s, are gettered to the dislocation core.

III. REQUIREMENTS OF A PASSIVATING BARRIER LAYER

To passivate a semiconductor both chemically and electrically, a barrier layer must cover semiconductor conformally, completely, and stably. It must be well bonded to the semiconductor and able to flex with it during processes which warp and flex the product. (These include oxidations, vacuum chucking for lithography, metallization, packaging and thermal transients at various times during production and use.) It must also block the diffusion of charge, atoms and ions.

To be stable chemically, a substance must have large activation barriers against reactions that would transform it to anything else. It must have a large cohesive energy. As elemental solids with large cohesive energies

also liberate large energies forming certain compounds, particularly nitrides and oxides, one wants a compound with a large heat of formation. In order to bond well to the semiconductor, the bond energy to those atoms must also be large.

In order to cover the semiconductor conformally and to flex with it without developing cracks or grain boundaries along crystalline planes, it is very advantageous that the barrier be amorphous, rather than crystalline. Thus, one wants an amorphous phase that is stable; it ought to have a sufficiently high glass transition temperature. It should also have a slow variation of viscosity with temperature around that point so that it can flow to conform continuously to the semiconductor without cracking, crystallizing, or imparting much strain to the semiconductor.

Directional covalent bonding is important to the stabilization of amorphous phases formed from a few elements with atoms of similar size (33). It allows for the formation of stable but non-crystalline bonding networks. Non-directional forces, i.e., ionic (e.g., NaCl), metallic (e.g., Cu) and van der Walls (e.g., solid Ar) binding, imply easy crystallization unless one has a mix of several different sized atoms. The only examples of stable amorphous solids formed without directional covalent bonding (known to this author) are metal alloys composed of several elements such that the atoms have distinctly different size. These atoms are very stable in their amorphous phase because they fill space much more efficiently than can atoms all of the same size (because metallic binding increases with electron density) and they would crystallize to different crystal structures only if they could diffuse enough to phase separate (34-37). Atomic diffusion is very slow through such alloys because there is so little free volume. These amorphous metals might be useful to stop atomic diffusion out of conductors, particularly if a better conductor like Cu or Au which must be kept out of the semiconductor were to be deposited on top of them, but they would not provide electrical isolation.

Thus, the requirement of chemical and phase stability together with a large band gap direct us to compounds with moderate ionicity and large heats of formation with strongly directional covalent bonds. To avoid phase separation, which would imply grain boundaries that trap charge and promote diffusion, there ought not to be more than one easily obtained phase. Thus the fact that the native oxides of GaAs include Ga_2O_3, $GaAsO_4$, and As_2O_5 implies that attempts to passivate GaAs with its native oxide must face this difficulty. Deposited layers, e.g., CVD a-SiO_2 layers, must face the problem of forming stable bonds between the SiO_2 and the GaAs despite a smaller heat of formation for such bonds. Also, when one deposits a layer rather than oxidizing the semiconductor, one looses the cleansing effect of the reaction; "dirt" tends to be burried rather than vaporized. Where one must deposit the layer, annealing will help to passivate the trapped "dirt".

The fact that Ge "snow plows" ahead of the oxide interface when Si-Ge alloys are oxidized so that the oxide is pure SiO_2 points up the importance of the difference in their heats of formation. (The enthalpy of formation at T = 25 C, ΔH_f^0, for a-GeO_2 is 5.57 eV compared with 9.36 eV for a-SiO_2.) Because a-SiO_2 has been demonstrated to be such a good material from which to for a passivating barrier layer, this snow plow effect lends much hope for Si-Ge heterostructure technologies that may overcome the limitations imposed by the failure of Si bipolar transistors to amplify below 135 K.

The Ge snow plow effect upon oxidation also shows that the dominant direction of Si diffusion at the oxidizing interface is from the bulk to

interface. This implies oxidation injects vacancies formed at the incoherent oxide interface into the bulk, not Si self-interstitials as some have claimed (38). Oxidation enhanced and retarded diffusion of dopants, and other atoms, is a complex and still unresolved problem (39). The implications for passivation are not yet clear; the fact that vacancies are quite mobile in Si and Ge at room temperature implies that they are likely to be gettered and present less trouble than in III-V compounds, where they are not so mobile on account of the formation of antisite defects for nearest neighbor hopping.

At this point we can see that we are being driven to the conclusion that thermal amorphous SiO_2 and Si_3N_4 are likely to be optimal materials from which to form passivating barrier layers. This should not be surprising in light of the experience of industry.

However, there remain the problems of the dangling bonds implied by the incoherent interface between a-SiO_2 and the semiconductor, as well as dangling bonds at vacancies. (One may define a vacancy in an amorphous covalent network as a site where bonds are left dangling.) As is well known the traditional remedy is to introduce H from H_2, H plasma, forming gas, H_2O steam, etc. in order to terminate these dangling bonds as H bonds. Because OH is an effective catalyst for the Si-O bond this also has the benefit of promoting compete coverage of the semiconductor by very thin layers and of reducing the strain between the layers due to differences in coefficients of thermal expansion.

On the other hand we are now aware of the problems that arise when H diffuses to dopants in the semiconductor and passivates them. It may be worthwhile to think of possible alternatives to H passivation of dangling bonds.

IV. Passivation with Hydrogen

H is extremely effective in passivating semiconductors. As it is light and small, it will rapidly diffuse through most solids. It will react with any dangling covalent bond to produce a "hydrogen bond." In most cases this reaction removes a deep level defect state associated with the dangling bond from the band gap of the semiconductor. In addition to this electronic effect, it also has the mechanical effect of filling up the free volume associated with the vacancy or other defect responsible for the dangling bond. In structural metals this leads to "hydrogen embrittlement"; here the important mechanical effect is to passivate atomic hopping processes, such as vacancy nearest neighbor hopping. Of course, it may take more than one H atom to passivate vacancy complexes and H_2 molecules may be present in the larger ones. We suppose this occurs for the EL2 defect, which we have good evidence is a family of defects (40) headed by an As_{Ga} together with a divacancy (24-26). It has been noted that EL2 is passivated by H and once passivated, more annealing at higher temperatures is required to reactivate the EL2 phenomenon than is required for reactivation of the dopants and several other deep level defects (5,6).

Unfortunately, we now know that H will also react with the acceptors and donors (that we must have to make devices) to remove their levels from the gap as well (41,42). This might have been anticipated from the fact that H reacts strongly with both electropositive elements (e.g., Li) and electronegative elements (F). H can also nucleate microsplits, very small cracks, in near surface regions by interposing two H atoms in each of the bonds across a cleavage plane to produce two H terminated surfaces (43).

It should also be noted that H is ubiquitous in semiconductor processing and is introduced, intentionally or not, from residual vacuum gasses, plasmas, annealing ambients, steam oxidations, acid etches, etc. at almost every step.

A few remarks about this vast subject seem to be called for partly because some recent results re the H passivation of DX have been claimed (6,44) to be supportive of the conclusion of the proposition that the corresponding reactions occur without dangling bonds. These claims seem in turn to stem from confusion re the energy of hydrogen bonds to various elements.

In 1986 Pearton et al. published (45) a plot of their empirical reactivation energies, E_D, for recovery of donors in GaAs from H passivation versus H to donor element bond energies. This plot is remarkably linear and accurate. This strict linearity is also remarkable because: a) The scales are not the same; the range of bond energies exceeds the range of E_D by about a factor 10 but the absolute magnitudes are of order 3 to 2.' No explanation in terms of screening would seem to suffice. b) The plot shows both H bond strength and E_D to be less for Si than for Ge. Experience with chemical trends of bond strengths, as well as the empirical literature, teach us that the H bond strengths to Si are actually larger than the comparable bond to Ge. c) The E_D values were deduced under the assumption that the attempt frequency, f, to break the H - donor bond is the same for all donors. The formula used was

$$E_D = kT \ln[(tf)^{-1} \ln(N_0//N)] \quad . \qquad (2)$$

In fact one might expect f to be greater for donors that bind the H more strongly, as the observed local vibrational mode frequencies are higher, and for lighter donors, which themselves vibrate at higher frequency in the lattice. Usually the lighter donor also binds H more strongly. Any trend of values for f in this direction will reduce the larger values of E_D deduced from (2) relative to the smaller ones. As it is, the range of E_D for f = const. (from 2.04 eV for Sn and Te to 2.16 eV for S) is barely larger than the claimed experimental error of (±) 0.04 eV.

However, the conclusion of Pearton et al. that the H is bonded to the passivated donor atoms, is supported by the fact that electron mobility increases upon passivation. If H acceptor were simply compensating the donors and distant from them, and if only ionized impurity scattering needs to be considered, the mobility should drop when H is added. Also, a rather stronger variation of E_D with H bond energy for dopants in Si is found.

Pearton et al. got their H - donor bond energies from a table in the 1975 *CRC Handbook of Chemistry and Physics* labeled "bond strengths of diatomic molecules." It suggests e.g., that Ge-H bond energy (quoted as 76.5 Kcal/mole) exceeds the Si-H bond energy (quoted as 74.6 Kcal/mole). This and similar tables are reproduced also in several other reference books; at least in the CRC Handbook it contains the caveats that some values were found for T = 0 K while others are for T = 25 C and that "The references have been chosen primarily as a key to the literature." Confusion also arises many values are in fact not obtained from diatomic molecules but from "single bond energies" deduced from larger molecules. Consideration of well established chemical trends, as well as some of the literature referenced in this table, show that in fact the energy is generally less for any particular Ge-H bond than for the corresponding Si-H bond. The present author relies on the tables published (46) by the U.S. National Bureau of Standards. These show e.g., for Si at T = 0 K the Si-H diatomic bond energy is 73.5 Kcal/mole and single bond energy for SiH_4, i.e., 1/4 of the energy

of the 4 tetrahedral bonds, which seems to be a more relevant quantity for the present problem, is 76.0 Kcal/mole and the corresponding single tetrahedral bond energy for Ge-H is 69.2 Kcal/mole. S.R. Gunn, one of the authors referenced in the CRC table, in fact concludes (47) very similar values for Si-H and Ge-H in the article published just before the particular one cited by CRC and gives an enlightening discussion of the problems of these determinations.

Table 1 gives some values for T = 0 K H bond energies that hopefully are appropriate for the present discussion. They are obtained from the NBS compilation. The molecule from which they were obtained is indicated also.

Given that the empirical evidence does imply that H does bind directly to donors (and to acceptors) in GaAs and that the choice of donor does affect E_D but to a degree much smaller than the range of molecular H-donor bond energies, we must conclude that the passivation/depassivation reactions are more complicated than simple making and breaking the donor-H bond. This might be anticipated from the fact that the positron annihilation experiment teach us that virtually all samples of GaAs contain large concentrations of vacancies, which is likely correlated to the fact that they are are also generally off perfect stoichiometry by parts per thousand. This, and the tendency of vacancies in compounds to hop to nearest neighbor sites thereby making antisite complexes, implies that virtually all deep levels will be complexes of related families (48) and will be difficult to getter in III-V's. Note that positrons (16) do not find vacancies in normal Si at room T; being an element, Si does not have a similar tendency to be non-stoichiometric. We noted that the variation of E_D with impurity is larger in Si than in GaAs.

It is evident that most of the non-stoichiometric defects in GaAs are present in neutral complexes. If the sample is grown Ga rich, then the non-stoichiometric defects include V_{As}, which is a deep donor and Ga_{As}, which is a double acceptor so, among others, the complex $V_{As}Ga_{As}V_{As}$ forms (22,23). If the sample is grown As rich, then V_{Ga} which is a deep acceptor tends to complex with As_{Ga}, which is a double donor, to form $V_{Ga}As_{Ga}V_{Ga}$. Note that the Ga rich complexes present Ga dangling bonds to the H at both the V_{As} and the Ga antisite, while the As rich samples present As dangling bonds at both point defects. Given that almost all GaAs sold to industry in recent years has been deliberately grown As rich to take advantage of the EL2 phenomenon to produce a semi-insulating substrate, we may assume that the samples used by Pearton et al. were also As rich so they present many As dangling bonds to diffusing H. Then the value of E_D is determined mainly by the energy difference between the As-H bond at the defect site in the sample, which is probably not much different from the 3.02 eV of the the AsH_3 molecule, and the energy of the H in its diffusion channel. There is evidence that the metastable channel site for the diffusing H in GaAs is a bond centered interstitial site (49); the same may be true of Si and other covalent semiconductors (50). Then we estimate the difference between the energy of the H atom in vacuum and in this channel site, i.e., the H atom adsorption energy for GaAs, is about 1.0 eV.

This picture may also explain why Pearton et al. estimate the activation energy for H diffusion in GaAs to be 0.5 eV (but increasing with doping) while Jalil et al. find (51) it to be 1.4 eV. We propose that the activation energy to hop from one interstitial site to another is of order 0.5 eV while long range diffusion is strongly affected by trapping at dangling bonds. If the bond is to a Ga atom, we estimate (Table 1) the detrapping energy to be 1.8 eV; if it is to an As atom we estimate 2.0 eV. Thus, the value found by Jalil et al. is assumed to be the thermal compromise between

Table 1. H Bond Energies at T = 0 K

Bond	ΔH (Kcal/mole)	ΔH (eV)	molecule
B–H	87.7	3.80	BH_3
C–H	98.1	4.25	CH_4
N–H	92.3	4.00	NH_3
O–H	101.4	4.39	OH
Al–H	67.1	2.91	AlH
Si–H	76.0	3.30	SiH_4
P–H	75.6	3.28	PH_3
S–H	86.8	3.76	SH_2
Ga–H	64.6	2.80	GaH
Ge–H	69.2	3.00	GeH_4
As–H	69.7	3.02	AsH_3
Se–H	74.7	3.24	SeH_2
Sn–H	59.2	2.57	SnH_4
Sb–H	60.4	2.62	SbH_3
Te–H	63.1	2.73	TeH_2

these detrapping values and the much smaller interstitial-interstitial hopping value.

We conclude that the rate of de-passivation of the donors by H is dominated by the large number of detrapping events of H from As dangling bonds that are required to remove the H from the vicinity of the donors and is only marginally affected by the single (or, at most, few) detrapping events of the H from the donor itself. Thus, the difference between the reactivation energy for these donors and for the DX center (which much evidence indicates (29,30) is a complex of a donor with an X = V_{As} that hops to nearest neighbor sites via reaction (1)), is expected to be only slight. The observation that this is true (6) is not evidence against the proposition that there is an X and indeed it is the V_{As}. Again, the fact that a much larger re-activation energy is found for EL2 is likely due to the fact that its divacancy cavity takes up more than one H before it is passivated and these form more stable molecular structures.

This is the sort of complicated defect interaction process that is best handled by Monte Carlo simulation methods. We have devised an efficient and inexpensive simulation program, VIDSIM, which can do (39,52,53) this and similar problems on microcomputers. (We supply VIDSIM to any one requesting it for the cost of the media.) Preliminary results of the simulation of vacancy hopping in III-V compounds indicates that real interactions are much more complicated than one might at first expect. This is indicated also by experiments on diffusion of Al in GaAs doped with different donors, which give distinctly different kinetics (54). We have concluded that this is due to the tendency of light atoms (e.g.,Al, Si and S) to engage in correlated hopping with vacancies and to promote the formation of multivacancy complexes.

V. PASSIVATION BY GETTERING

Gettering is the deliberate trapping of impurity atoms, vacancies or host interstitials at some site where they will do no harm.

One can accomplish gettering by binding the point defects via a chemical reaction, an elastic force field, an electrostatic force field, or a magnetic field. Chemical gettering is widely used in vacuum systems but has not been much used in solid state devices. Magnetic getting is used for some machines, but magnetic forces are usually too weak to be useful for solid state devices. The recovery of CdS–Cu_2S solar cells by placing them in the dark, so that the junction field will be maximal, is an example of the successful use of electrostatic fields to getter point defect (Cu interstitials) that degrade the performance of a solid state device (55). However, the most common and important examples of gettering in the semiconductor industry use elastic fields to bind the point defects. The single most important example is use of SiO_2 precipitates in Czochralski Si substrates for "internal gettering". These precipitates punch out dislocations and otherwise strain the lattice to trap harmful point defects away from the device layers nearer the surface, which must be denuded of O.

It is well known that dislocations from other sources, e.g., misfit stress between layers of different doping or heterojunctions, will also getter point defects in their elastic strain fields. Dislocations will also incorporate point defects into their core structure when they climb. It is also well established that dislocations can me made to glide in most practical semiconductors by laser excitation above a critical intensity (56) and that below this critical intensity they will climb by gettering point defects from the surrounding material at a rate that is dramatically enhanced by the electron–hole recombination enhanced diffusion of those point defects. We have already noted that the X of the DX center is one of the point defects that may be gettered in this way (32). This suggests the use of ordinary device lithography to define regions in the product into which one wants dislocations to glide under laser excitation so that they will subsequently passivate nearby device regions by gettering out point defects, for example by recombination enhanced diffusion under a lower level of laser excitation. We are presently engaged in exploring the practical applications of this "optical glide and climb" passivation to III–V heterostructure devices which contain many threading dislocations due to large misfit stresses.

The author expects many advances in passivation technology through the use of various forms of gettering in the future.

VI. PASSIVATION OF DANGLING BONDS WITHOUT H, e.g., WITH Ar

When we consider alternatives to H to passivate dangling bonds, it seems worthwhile to consider elements at the opposite extreme of chemical reactivity from H. Whereas H is the most reactive element, the noble gases are the least reactive. As does H, the noble gas atoms have the attractive feature that they generally diffuse rapidly in most semiconductor lattices. They also have the feature that they precipitate into vacancies, and other cavities, where they mechanically block further atomic migration and seem even to bond chemically into the host lattice (57).

Noble gases can easily be ion implanted into semiconductor products, incorporated from plasmas, or introduced from the amorphous metals (37) that were suggested as diffusion stops against Cu and other optimal conductors. The

use of noble gases for vacancy complex passivation may be particularly useful in avoiding problems of H passivation in an intense irradiation environment.

VII. CONCLUSIONS

This author is convinced that the best help a theorist can give to the main stream of semiconductor passivation technology is via Monte Carlo computer experiments, using e.g., VIDSIM (39,52,53). The problems of H passivation of deep level defects of current practical interest are simply too complex to be resolved by *ab initio* analytic techniques. As for new innovations, it seems that dislocations ought to be considered an opportunity rather than a lattice problem. They are almost inevitable in the strained layer super lattice structures currently being explored; they can be made to glide or to climb rapidly by laser irradiation that can be be defined lithographically. The gettering of very problematic point defects, like the X of the DX center, by dislocation climb has been demonstrated. The use of amorphous metals as a diffusion stop may find application, as may the use of noble gas atoms to bind up dangling bonds in vacancy complexes.

REFERENCES

1. J. Dabrowski and M. Scheffler, *Phys. Rev. Lett.* 60, 2183 (1988) and therein.
2. D.J. Chadi and K.J. Chang, *Phys. Rev. Lett.* 61, 873 (1988) and therein.
3. M. Mizuta, M. Tachikawa, H. Kukimoto, and S. Minomura, *Jpn. J. Appl. Phys.* 24, L143 (1985).
4. H.J. von Bardeleben, D. Stievenard, D. Deresmes, A. Huber, and J. C. Bourgoin, *Phys. Rev.* B 34, 7192 (1986) and therein.
5. J. Lagowski, M. Kaminska, J. M. Parsey, H. C. Gatos, M. Lictensteiger, *Appl. Phys. Lett.* 41, 1078 (1982).
6. J.C. Nabity, M. Stavola, J. Lopata, W.C. Dautermont-Smith, C.W. Tu, and S.J. Pearton, *Appl. Phys. Lett.* 50, 921 (1987).
7. W. Shockley, *Electrons and Holes in Semiconductors* (Van Norstrand, New York, 1950) esp. chapt. 2.
8. J.C. Phillips, *Bonds and Bands in Semiconductors* (Academic Press, New York, 1973).
9. J.A. Van Vechten in *Handbook on Semiconductors*, Vol. 3, ed. by S.P. Keller (North-Holland, Amsterdam, 1980) chapt. 1.
10. J.A. Van Vechten and R.M. Martin, *Phys. Rev. Lett.* 28, 446 (1972).
11. W. Hanke and L.J. Sham, *Phys. Rev. Lett.* 33, 582 (1974).
12. G. Strinati, H.J. Mattausch and W. Hanke, *Phys. Rev. Lett.* 45, 290 (1980).
13. W. Hanke, G. Strinati and H.J. Mattausch in *Recent Developments in Condensed Matter Physics*, ed. by J.T. Devreese (Plenum, New York, 1981) Vol. 1.
14. R.M. Martin, Festkorperproblem, *Advan. in Solid State Phys.*, XXV, 3 (1985).
15. Y-T. Shen, B.M. Bylander and L. Kleinman, *Phys. Rev.* B 36, 3465 (1987).
16. S. Dannefaer and D. Kerr, *J. Appl. Phys.* 60, 1313 (1986).
17. S. Dannefaer, P Mascher and D. Kerr, *J. Phys. Condens. Matter.* 1 3213 (1989) and therein.
18. M.J. Puska, O. Jepsen, O. Gunnarson and R.M. Nieminen, *Phys. Rev.* B 34, 2695 (1986).
19. S. Dannefaer, private communication.
20. H.J. Lim, H.J. von Bardeleben, and J.C. Bourgoin, *J. Appl. Phys.* 62, 2738 (1987).

21. A.S. Jordan, A.R. Von Neida, R. Caruso and C.K. Kim, *J. Electrochem. Soc.* 121, 153 (1974).
22. J.A. Van Vechten, *J. Electrochem. Soc.* 122, 419 + 423 (1975).
23. J.B. Van der Sande and E.T. Peters, *J. Appl. Phys.* 45, 1298 (1974) and *J. Appl. Phys.* 46, 3689 (1975).
24. J.F. Wager and J.A. Van Vechten, *Phys. Rev.* B 35, 2330 (1987) and *Phys. Rev.* B 39, 1967 (1989).
25. Zou Yuanxi, *Mater. Lett.* 5, 203 (1987).
26. Wang G-Y., Zou Y-X., S. Benakki, A. Goltzene, and C. Schwab, *Phys. Rev.* B 38, 10953 (1988).
27. K. Bonde-Nielsen, H. Grunn, H. Haas, F.T. Pedersen and G. Weyer, *J. Electron. Mater.* 14a, 1065 (1985).
28. P. Gibart, D.L. Williamson, B. El Jani, and P. Basmaji, *Phys. Rev.* B 38, 1885 (1988) and therein.
29. J.A. Van Vechten, *J. Phys: Condens. Matter* 1, 5171 (1989).
30. D.L. Williamson and P. Gibart in *The DX Center in III-V and Ternary Compounds* ed. by J. C. Bourgoin (Trans-Tech, Switz. 1990).
31. J.A. Van Vechten, *Mater. Res. Soc. Symp. Proc.* 46, 83 (1985).
32. D.V. Lang, P.M. Petroff, R.A. Logan and W.D. Johnston, *Phys. Rev. Lett.* 42, 1353 (1979).
33. J.C. Phillips and M.F. Thorpe, *Solid State. Commun.* 53, 699 (1985) and therein.
34. G.S. Cargill, *J. Appl. Phys.* 41, 2249 (1970).
35. P. Chaudhari, J.J. Cuomo, and R.J. Gambino, *IBM J. Res. Devel.* 17, 66 (1973).
36. P. Chaudhari and D. Turnbull, *Sci. Amer.* 242, 1980 (1980).
37. J.A. Van Vechten, R.J. Gambino and J.J. Cuomo, *IBM J. Res. Devel.* 23, 278 (1979).
38. P.M. Fahey, P.B. Griffin, and J.D. Plummer, *Rev. Mod. Phys.* 61, 289 (1989) and therein.
39. J.A. Van Vechten, U. Schmid, and N.C. Myers, Yokohama Conf. on Defect Control in Semiconductors, 1989, to be published.
40. M. Taniguchi and T. Ikoma, *Appl. Phys. Lett.* 45, 69 (1984).
41. C.T. Sah in *Semiconductor Si, EMIS Datareview* 4 (INSPEC, London, 1988) pp. 584 and therein.
42. W.C. Dautremont-Smith, *Mater. Res. Soc. Symp. Proc.* 104, 313 (1988) and therein.
43. N.M. Johnson, F.A. Ponce, R.A. Street and R.J. Nemanich, *Phys. Rev* B 35, 4166 (1987).
44. R. Mostefaoui, J. Chevallier, A. Jalil, J.C. Pesant, C.W. Tu, and R.F. Kopf, *J. Appl. Phys.* 64, 207 (1988).
45. S.J. Pearton, W.C. Dautremont-Smith, J. Chevallier, C.W. Tu, and K.D Cummings, *J. Appl. Phys.* 59, 2821 (1986).
46. D.D. Wagman, W.H. Evans, V.B. Parker, I. Halow, S.M. Bailey and R.H. Schumm, NBS Tech. Note 270-3 (National Bureau of Standards, Washington, 1968).
47. S.R. Gunn and L.G. Green, *J. Phys. Chem.* 68, 946 (1964).
48. J.A. Van Vechten, *J. Phys. C: Solid State Phys.* 17 L933 (1984).
49. R.F. Kiefl and T.L. Estle in *Hydrogen in Semiconductors* ed. J.L. Pankove and N.M. Johnson (Academic Press, New York, 1990) and therein
50. C.G. Van de Walle, Y. Bar-Yam, and S.T. Pantelides, *Phys. Rev.* B 60, 2761 (1988).
51. A. Jalil, C.H. Gratteain, C.L. Grattepain, A. Huber and Chevallier, to be published.
52. J.A. Van Vechten and U. Schmid, *J. Vac. Sci. Technol.* B 7, 827 (1989)
53. U. Schmid, N.C. Myers, and J.A. Van Vechten, *Comp. Phys. Commun.* (in press).

54. P. Mei, S.A. Schwarz, T. Venkatesan, C.L. Schwartz, J.P. Harbison, L. Florez, N.D. Theodore and C.B. Carter, *Appl. Phys. Lett.* 53, 2650 (1988).
55. J.A. Van Vechten in *CdS Solar Cells and other Abrupt Heterojunctions* ed. by K.W. Broer and J.D. Meakin (U. of Delaware, Newark, 1975) pp. 191.
56. B. Monemar, R.M. Potemski, M.B. Small, J.A. Van Vechten and G.R. Woolhouse, *Phys. Rev. Lett.* 41, 260 (1978) and therein.
57. V.D. Tkachev, A.V. Mudryi and N.S. Menrev, *Phys. Stat. Sol.* A 81, 373 (1981) and therein.

Defeating Compensation in Wide Gap Semiconductors by Growing in H that is Removed by Low Temperature De-Ionizing Radiation

James A. Van Vechten,[1]* James David Zook,[2] Robert D. Horning[2] and Barbara Goldenberg[2]

*Supported in part by U.S. AFOSR-89-0309.

[1]Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331-3211 USA

[2]Sensor and System Development Center, Honeywell Inc., Bloomington, Minnesota 55420 USA

(Received

ABSTRACT

We propose a general method to obtain high conductivity of either type in wide gap semiconductors where compensation by native point defects normally limits conductivity of one or both types. We suggest that the successes of Amano et al. and of Nakamura et al. in obtaining more than 10^{18} cm^{-3} holes in GaN are particular examples of the general process that we propose.

KEYWORDS: hydrogen, compensation, GaN, conductivity, wide gap semiconductors.

Attempts to obtain practical levels of n-type or of p-type conductivity in wide band gap semiconductors are usually frustrated by the tendency of the material to self-compensate the shallow donor or shallow acceptor dopant impurities, that one deliberately introduced into the material. The spontaneously generated native point defects that do the self-compensation might be vacancies, antisite defects, host interstitials or complexes among these. The root cause of this tendency towards self compensation is the natural tendency to minimize free energy.[1] An important fraction of the total free energy of the sample is minimized when the Fermi level, E_f, is at the intrinsic level, E_i, of a semiconductor or insulator. This is because occupancy of the bonding states of the valence band, which are pushed down relative to the free electron gas levels of the jellium model by the crystal potential, reduces the enthalpy by the same amount that occupancy of the corresponding anti-bonding states of the conduction band would raise the enthalpy.[1-5] Thus, if by doping one were to move E_f so close to one or the other band edge that nature could reduce the free energy of the sample by generating some collection of native defects with appropriate ionization levels in the band gap and letting the free carriers fall from the Fermi Sea into these levels, then nature will do this at a rate limited by the activation barriers for the defect generation and diffusion processes at the ambient temperature. If this rate is rapid enough, one's attempt to obtain practical conductivity will be frustrated.

Several methods to circumvent the tendency to self-compensation have been developed. A well-known example is the use of ion implantation to introduce the dopant impurities at concentrations far beyond equilibrium thermodynamic limits. However, one usually has to anneal out the damage concomitant with ion implantation and compensating defects are often introduced in that process. It has been found that laser annealing or rapid thermal annealing processes, which cause a major splitting of the quasi-Fermi levels while the damage is being removed, can produce higher levels of electrical activation (i.e., less self-compensation) than ordinary thermal annealing.[6,7] Rapid quenching and crystallization can also be used to obtain a crystal with active doping levels far from thermal equilibrium, which one may be able to retain long enough for practical purposes by keeping the sample cool after preparation. Unfortunately, the processes required to metallize, mount and package

semiconductor devices are often incompatible with a requirement that it be kept so cool that self-compensation does not occur.

Other strategies for circumventing self-compensation involve nuclear transmutation. One can arrange that the shallow dopants be created by nuclear decay only after the device has been fabricated and packaged. Alternatively, one can deliberately compensate the intended shallow dopant impurities with radioactive impurities which decay to neutral host or impurity atoms with times long compared with the fabrication process.[8] In either case E_f moves to the point that self-compensation would be energetically favorable only after high temperature processing has been completed. Such methods are limited by the availability of useful isotopes, problems with residual radioactivity, and with the fact that the recoil energy in a nuclear transformation reaction is often large enough to displace the radioactive atom from the lattice site and create a vacancy interstitial pair, which may compensate the sample.

We here note another alternative to circumvent self-compensation. The strategy is to introduce H ions to compensate the intended shallow dopant impurities during crystal growth and subsequently to remove them by converting them to neutral H atoms and diffusing them out of the active region to a free surface or to an internal void. On the free surface or the void surface, the neutral H atoms combine to form H_2 molecules and desorb as the elemental gas. We make use of the facts that atomic H generally produces both deep donor and deep acceptor ionization levels[9-11] in Si and in wider band gap semiconductors and that, while these H^+ and H^- ions tend to cluster with the shallow acceptor impurities or shallow donor impurities respectively, the neutral H diffuses easily in the host lattice and migrates large distance before re-ionizing. We also make use of the fact that, when E_f is kept near E_i, there is no thermodynamic driving force to generate more native defects; the native defects are not needed to compensate the shallow dopants because that has already been accomplished with the H ions. To neutralize the H ions, we introduce "de-ionizing" radiation, i.e., radiation that produces electron-hole pairs in the desired region. This can be done with above band gap light or with electron beams or, perhaps, directly by direct injection. The major considerations in designing the de-ionization process are that it should neither heat nor damage the host lattice but should provide adequate currents for adequate times in the desired

region so that the H ions will be neutralized enough times to wander to the intended sinks. One must prevent native defect reactions from occurring, because they would tend to reintroduce compensation. This precludes electron beam energies larger than 100 keV and fluxes that would raise lattice temperatures to more than about 300 C. Note that the range of the effect is not the primary range of the incident beam of electrons or light but the depth to which the electron-hole pairs that the beam produces diffuse.

A practical method to provide the H ions required to compensate the doping during crystal growth is to grow in the presence of a H plasma. Common methods of chemical vapor deposition, CVD, or organometallic chemical vapor deposition, OMCVD, crystal growth can easily accommodate this requirement for "plasma enhancement". CVD and OMCVD can also supply H ions by cracking H bearing molecules on the crystal surface. One can imagine methods to incorporate H ions in hydrothermal or flux growth systems and for liquid phase growth.

We must now note that at least two reports[12,13] have already been published which we believe demonstrate the efficacy of the general method just described for the particular case of p-type GaN. Amano et al. and Nakamura et al. do not discuss how the process which they call "low energy electron beam irradiation" or LEEBI may have succeeded in producing samples with hole concentrations, $p > 10^{18}$ cm^{-3} in this wide gap semiconductor. Previous reports had concluded that, due to self-compensation, GaN can be made only insulating or n-type. The native defects responsible for this are thought to be N vacancies or N on Ga site antisite defects.[1,4,8,14]

The LEEBI process described by Amano et al. and by Nakamura et al. appears to be consistent with the prescription given above for the general method. The samples of GaN were grown by OMCVD with H present. As grown, they were semi-insulating or n-type. We posit that the as grown material is insulating because most of the shallow acceptors, A^-, are compensated by H^+ ions that reside at interstitial sites in the same unit cell as the A^- substitutional site. Where there was n-type conductivity, we posit that it was the result of insufficient stoichiometry control. We further posit that the samples have, given the large concentrations of shallow acceptor impurities that were incorporated, relatively low concentrations of vacancies, antisite defects, and host interstitials

because the compensation was accomplished with the H^+ interstitials. The samples were irradiated with an electron beam of 5 to 15 keV incident energy for several hours, which caused them to become p-type with good conductivity. The 10 keV initial kinetic energy of incident electrons is not enough to displace atoms from bonded sites in the host lattice,[1,8] and the flux is low enough that there is no substantial heating so that no new vacancies, interstitials or antisite defects can be introduced. However, the irradiation does produce electron-hole pairs. Both carriers and their excitons tend to collect at the H^+ - A^- donor acceptor pairs. These electrons neutralize the H^+ interstitials; the neutral H atoms diffuse rapidly as interstitials until they either re-ionize and rebind to an A^- or reach a surface and combine with another H atom to form the gas. Because the only effect is that the H leaves, this leaves the sample with a large concentration of uncompensated substitutional shallow acceptors trapped at low temperature in a GaN host lattice that is relatively free of native point defects and now denuded of H. Even though this is not the state of lowest free energy, the configuration is metastable with large activation barriers against re-establishing the favored insulating state. (Recall that diamond is not the stable phase of C at normal temperature and pressure, but it lasts "forever".)

We conclude that the general process we have described for doping semiconductors to levels heretofore unattainable by first growing the material heavily compensated by H ions and then removing the H ions with a low energy, modest flux beam of de-ionizing radiation accounts for the case of p-type GaN, as has been reported[12,13] by Amano et al. and by Nakamura et al. Furthermore, we posit that similar processes will produce p-type conductivity in other wide gap semiconductors, including BN, AlN, BeO, ZnO, ZnS, and CdS where self-compensation of acceptor doping has been limiting. For the case of ZnSe, analysis of native defect equilibria has long predicted[1] that p-type doping can be obtained by near equilibrium processes. Modest p-type conductivity has been reported for many years with much better conductivity found recently with the advent of N doping. Some of the recent success has been had with processes where there is no obvious source of H. We suggest that, while ZnSe can be doped p-type without our proposed H process, one might improve the p-type conductivity with the H process because it is expected to increase the solubility of acceptors and to decrease the solubility of any donor impurities as well as the concentration of native donors. Furthermore, we propose that

the analogous process with H^- ions, that compensate shallow donors and can be neutralized by injecting holes, can produce useful enhancement of n-type conductivity in cases where n-type conductivity is limited by self-compensation.

To practice the method, one deliberately incorporates during crystal growth not only the desired shallow acceptor or donor dopant impurities but also large quantities of H. The material thus grown is insulating due to the compensation of the shallow dopant levels by the H deep donor or deep acceptor levels. As this compensation is provided by the H, the sample does not respond to the doping by generating native defects. Once the sample is cool and defect creation would necessarily be very slow, one can remove the H by first converting the + or - ions to neutral H with radiation that produces electron-hole pairs. The neutral H will diffuse long distances and will eventually reach the sample surface or internal voids. When pairs of neutral H atoms meet on a free surface, they will desorb as the gas. The shallow dopants are left uncompensated in the wide gap host which can not respond in practical times.

References

1) J.A. Van Vechten: Handbook on Semiconductors, Vol. 3, edited by S. P. Keller (North Holland, Amsterdam, 1980) Chap. 1 and therein.

2) V. Heine and J.A. Van Vechten, Phys. Rev. B 13, 1622 (1976).

3) F.A. Kröger, Ann. Rev. in Mater. Sci. 7, 449 (1977).

4) U.V. Desnica, N.B. Urli, and B. Etlinger, Phys. Rev. B. 15, 4119 (1977).

5) J.A. Van Vechten, Mater. Res. Soc. Symp. Proc. 46, 83 (1985).

6) J.A. Van Vechten, R. Tsu, and F.W. Saris, Phys. Lett. A 74, 422 (1979) and therein.

7) A. Kamgar, F.A. Baiocchi, and T.T. Sheng, Appl. Phys. Lett. 48, 1090 (1986) and therein.

8) U.V. Desnica, Phys. Stat. Solidi A 39, K33 (1977) and therein.

9) C.T. Sah in Semiconductor Si, EMIS Datareview 4 (INSPEC, London, 1988) and therein.

10) J.I. Pankove, in *Hydrogen in Semiconductors* edited by J.I. Pankove and N.M. Johnson (Academic Press, San Diego, 1991) and therein.

11) J. Chevallier, B. Clerjaud and B. Pajot, ibid.

12) H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, Jpn. J. Appl. Phys. 28, L2172 (1989).

13) S. Nakamura, M. Senoh, and T. Mukai, Jpn. J. Appl. 30, L1708 (1991).

14) M. Sano and M. Aoki, Jpn. J. Appl. Phys. 15, 1943 (1976).

J. Phys.: Condens. Matter 2 (1990) 281–293. Printed in the UK

The temperature dependence of band offsets for semiconductor heterojunctions in general and for the particular cases of AlAs–GaAs and HgTe–CdTe

J A Van Vechten† and K J Malloy‡

† Center for Advanced Materials Research, Department of Electrical and Computer Engineering, Oregon State University, Corvallis, OR 97331-3211, USA

‡ US Air Force Office of Scientific Research, Directorate of Electronic and Material Sciences, Bolling Air Force Base, DC 20332, USA

Received 17 July 1989

Abstract. We devise a simple theory for the temperature dependence of the valence and conduction band offsets in semiconductor heterojunctions using the thermodynamic point of view. The temperature dependencies of the offsets originate from the separate contributions of holes in the valence bands and of electrons in the conduction bands of the two semiconductors to the temperature dependence of their respective band gaps. We use the earlier determination of these contributions by Heine and Henry from isotope shifts of luminescent lines due to impurities. By considering this temperature dependence of the band offset we suggest an explanation for the discrepancy between the determinations of AlAs–GaAs valence band offset by Wolford *et al* and by Batey and Wright. Whereas for most pairs of semiconductors the bands move in the same direction with varying temperature, for the particular case of HgTe–CdTe they move in opposite directions. From this we predict a much greater than usual temperature dependence for the band offsets for HgTe–CdTe junctions and reconcile the major discrepancy between valence band offsets determined by Kowalczyk *et al* and by Chow *et al*.

1. Introduction

With the dramatic advance of semiconductor heterojunction technology [1–10] has come a concomitant increase of interest in the origin [11–18] of the valence and conduction band offsets, or discontinuities, ΔE_v and ΔE_c, respectively. In spite of the implied simiplicity in the concept of band offsets, no single theoretical description has been acknowledged as correct. To add an additional factor to this uncertainty, we wish to explore the temperature dependence of the band offsets, an issue ignored until raised by recent experiments [19, 20].

Approaches to the problem of the band offsets divide into two categories: thermodynamic, and electronic structure calculations. In addition to differences in the method of calculation between the two approaches, there is a difference in the way they view the small, but detectable, effect of depolarising fields set up around the boundaries of the samples as a result of dipole layers that are concomitant with those boundaries. The distinction is particularly clear and acute for the case of the band offset to vacuum, i.e. for the questions of the 'work functions' or 'ionisation potentials' of metals or of

0953-8984/90/020281 + 13 $03.50

semiconductors. For the purposes of the present paper, which is concerned with the temperature dependence of the band offsets, the point to keep in mind is that the effect of depolarising fields on junctions between two semiconductors is small at all temperatures and, owing to its origin, should be expected to have very little temperature dependence in the range $0 < T < 310$ K in which the data we discuss were taken. However, there seems to be such a semantic disagreement, and some confusion, between the two camps over this issue that some space must be taken to discuss it. ([11] and [13] contain previous attempts to resolve these semantic arguments and confusion.)

Those who use thermodynamic approaches are usually concerned with differences in free energy, chemical potential etc. between bulk phases. For such a consideration, any junction between two phases is irrelevant; there need not be any junction so long as they can somehow exchange the energy, particles etc. relevant to the problem. Thus, with the thermodynamic approach to the band offset (and work function) problems [11–13, 15], one ascribes absolute values (usually relative to an idealised 'vacuum level' [11]) to the enthalpies, entropies, and standard free energies of the valence and conduction band edge density of states distributions in the equipotential bulk regions of each semiconductor [21, 22], when no bias field is applied externally. (Recall that the standard free energies are the total free energies minus all explicitly concentration dependent entropy terms; in particular, the entropy terms resulting from the statistical distribution of the free carriers among the band states are absent from these standard free energies. Thurmond has given a particularly clear and authoritative account [21] of this point. Any good text book contains a general discussion, usually when treating the law of mass action.) Those who use this thermodynamic approach then define the band offset as simply the difference between these absolute standard free energies:

$$\Delta E_v(\mathrm{A/B}) = E_v(\mathrm{A}) - E_v(\mathrm{B}) \tag{1}$$

$$\Delta E_c(\mathrm{A/B}) = E_c(\mathrm{A}) - E_c(\mathrm{B}). \tag{2}$$

The result is, by this definition, independent of crystallographic orientation and transitive from one material to another; that is, for example,

$$\Delta E_c(\mathrm{AlAs/GaAs(110)}) = \Delta E_c(\mathrm{AlAs/GaAs(100)}) \tag{3}$$

$$\Delta E_v(\mathrm{AlAs/GaP}) = \Delta E_v(\mathrm{AlAs/GaAs}) + \Delta E_v(\mathrm{GaAs/GaP}). \tag{4}$$

Furthermore, for any pair of semiconductors A and B,

$$\Delta E_v(\mathrm{A/B}) - \Delta E_c(\mathrm{A/B}) = \Delta E_{cv}(\mathrm{B}) - \Delta E_{cv}(\mathrm{A}) \tag{5}$$

where

$$\Delta E_{cv} = E_c - E_v \tag{6}$$

denotes the band gap, E_v is the absolute standard free energy of the valence band edge distribution, and E_c is the absolute standard free energy of the conduction band edge distribution.

The fact that, by the thermodynamic approach definition, band offsets must be independent of orientation and transitive can readily be appreciated by considering, as in figure 1, the addition of a third material, C, to form junctions to the first two, A and B. The change in energy, enthalpy, total free energy or standard free energy upon transferring one carrier from the bulk of A to the bulk of B must be independent of the path utilised. Thus, the transit may be directly through the A/B junction or through the A/C junction followed by a path through sample C and through the C/B junction.

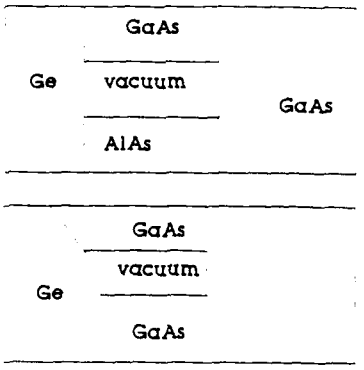


Figure 1. The change in energy etc when an electron goes from a particular state at a particular point in the bulk of Ge to a particular state at a particular point in the bulk of GaAs cannot depend upon the path taken. Thus, the bulk-to-bulk band offset, the definition used by those who use the thermodynamic approach, cannot depend on the orientation of the junction and must be a transitive property.

Because this is true regardless of the orientation of either junction or of the material C, which might be vacuum, it is plain that these bulk-to-bulk band offsets must be transitive and independent of orientation.

Now, it is well known [11] that when one measures the quantity generally known as the 'work function' of a metallic or semiconducting crystal (e.g. by photoemission), the result depends upon crystal orientation to a small degree; it depends very sensitively on the presence of any foreign atoms, e.g. Cs or O, on the surface. It is also well known [11] that the reason for this is that such experiments do not measure the chemical potential difference etc between the bulk of the crystal and the vacuum level, but instead they measure an effective potential energy barrier between a region in the sample near the surface and a region close above the sample, and because the termination of the sample at surfaces of various orientation, chemistry and condition produces electrostatic dipoles that depend upon that orientation, chemistry and condition. These dipoles produce depolarising fields that extend around the exterior of the sample, and, to a much smaller degree, extend also within the sample near the surfaces (or junctions). These fields would be irrelevant to the experiment if the electron were making a transition from the bulk to the idealised point at infinite separation from the sample and from all other charges, but practical experiments do not work that way

Those who use the thermodynamic approach realise, and readily admit, that practical experiments that measure the 'work function' do find a quantity that is neither transitive nor orientation independent. They distinguish [11–13] the 'work function' from the thermodynamic parameter that they would call the 'band offset between the sample material and vacuum'. Some attempt [13] to avoid confusion on this point is made by calling the thermodynamic 'band offset to vacuum' the 'ionisation potential'. It is quite a tractable problem to extract [11] the 'ionisation potential' from an adequate set of data on the 'work function'.

Unless the sample is heated so much that its surface chemistry changes or the surface atoms diffuse enough to change their local surface array, the dipole layer fields that produce this distinction between 'work function' and 'ionisation potential' will not change. If the surface dipoles do not change with T, then the magnitude of the distinction

will not change with T. Thus, the dipole fields will then make no contribution to any measured dependence on T of either the 'work function' or the 'ionisation potential'.

In general, practical junctions between crystalline samples also produce electrostatic dipoles that also produce fringing, or depolarisation fields. For semiconducting samples these fields are much weaker than those in vacuum because of the conductivity of the two solids. However, they do have a finite effect upon the effective barrier a carrier encounters transiting the junction in a practical experiment. Just as the chemistry of the surface has much more effect on the 'work function' that does the crystal orientation, the effective barrier at a semiconductor junction is expected to be much more sensitive to the local chemistry of the interface than to its orientation. (Indeed, the Freeouf–Woodall and Spicer defect models for Schottky barrier heights attribute a dramatic effect to a local chemistry at the interface different from that of the bulk sample.) Again, those with a thermodynamic point of view regard these junction field effects as a minor nuisance that should be extracted from the raw data in order to determine the true, bulk-to-bulk band offset, which must be transitive and orientation independent.

With the electronic structure approach [14, 16–18], one solves a Schrödinger equation for the particular chemistry and orientation of the junction as exactly and as rigorously as possible. With such an approach one has to deal explicitly with the junction dipole in detail. Those who use this approach regard these dipoles and their depolarisation fields with more respect than do those who use the thermodynamic approach. They generally include their effect when they seek to relate the results of their Schrödinger equation calculations, which are eigenvalues and not free energies, on the two sides to ΔE_v and ΔE_c. (The problem of relating eigenvalues to the enthalpies and free energies of the band density of states distributions is discussed on pages 8 to 18 of [22].)

Thus, with the electronic structure approach, one includes the small effect of the depolarising field in one's definition of the 'band offset' which is then an effective barrier height for transit of the carrier. Fortunately, the two approaches are now beginning to concur [18] with one another in cases like AlAs/GaAs where the close matching of the lattice constants makes for very simple junctions.

The effective barrier definition of the 'band offset' depends on the sample geometry and the distance of the two reference points from the interface. In general, it can be non-transitive and orientation dependent. It may be claimed to be more directly related to the result of practical experiments, but, in our view, it is not nearly so well nor clearly defined a concept and it is not nearly so easily applied to the question of the temperature dependence of the band offsets. Again, the reader should keep in mind that, because these junction dipole fields should not depend on T in the range of the relevant experiments, this distinction between the definitions of 'band offset' will have no significance for any discussion of the *temperature dependence* of the band offset.

The purpose of the present work is to address the problem of the temperature dependence of the band offsets. Although it seems not to have been done before, we will see that this is a very simple problem if one uses the thermodynamic approach. (It would appear to be rather difficult from the electronic structure approach.) We see from equations (1) to (6) that all that is required is a determination of the variation of E_v and E_c with T for the two semiconductors that form the junction. Most of what is required to establish $E_v(T)$ and $E_c(T)$ has already been developed and published with reference to the closely related problem of $\Delta E_{cv}(T)$, for which empirical values are established for most semiconductors [21].

Also for the problem of $d\Delta E_{cv}/dT$ there is a simple thermodynamic approach [22, 23] which avoids the difficulties of electronic structure calculations. One notes that the

explicit effect of T on the electronic eigenvalues is a very small part of the total $d\Delta E_{cv}/dT$ (as can be determined from the coefficient of thermal expansion and the pressure dependence of ΔE_{cv}) and the vast majority of the effect comes from the electron–phonon interaction. The electron–phonon interaction can be evaluated by considering the small perturbative effect upon the phonons of the thermally excited electrons, the e_c, and holes, the h_v, which are never more [21] than 1 part in 10^4 of the bonding electron density, using a simple bond-charge theory [24, 25] of phonon energies. This approach has been shown to give quantitative agreement with experiment for the phonon frequencies [26], as well as for $\Delta E_{cv}(T)$, up to the melting points of Si and Ge. (Of course, there is also an electronic structure approach [27] which considers the massive perturbation of (vibronic) electronic levels by phonons, which are several times more numerous than are host atoms. These two approaches now concur rather well.)

2. Temperature dependence of band offset

With the thermodynamic approach [21, 22] one readily sees that for most semiconductors ΔE_{cv} decreases with increasing T because the thermal excitation of both the e_c and h_v softens transverse acoustic phonon modes by (respectively) putting charge into antibonding states in the conduction band and taking it away from bonding states in the valence band. This results in a large positive (standard) entropy, ΔS_{cv}, for the reaction [21, 22] that thermally excites the e_c and h_v,

$$0 \leftrightarrow e_c + h_v \qquad \Delta E_{cv}. \tag{7}$$

As ΔE_{cv} is the (standard) chemical potential for $e_c + h_v$ pairs, it is equal to a free energy and follows the universal relation [21] of Gibbs

$$\Delta E_{cv}(T) = \Delta H_{cv}(T) - T\,\Delta S_{cv}(T) \tag{8}$$

where ΔH_{cv} is the corresponding (standard) enthalpy of the reaction. $\Delta E_{cv}(T)$, $\Delta H_{cv}(T)$, and $\Delta S_{cv}(T)$, as determined by Thurmond in [21] for the case of GaAs are plotted in figure 2. Note that basic considerations of thermodynamics require that $\Delta S_{cv}(T = 0\ \mathrm{K}) = 0$, and if $\Delta S_{cv} > 0$ for $T > 0$, then $d\,\Delta E_{cv}(T)/dT < 0$ and $d\,\Delta H_{cv}(T)/dT > 0$.

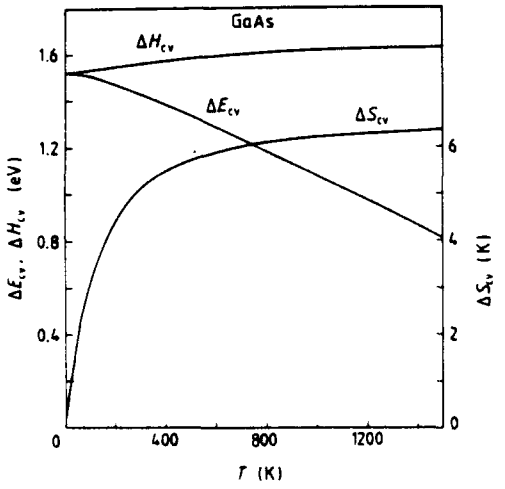


Figure 2. The empirical variation with temperature, as determined by Thurmond [21], of the free energy, the enthalpy, and the entropy of the band gap, i.e. of the creation of a pair of free carriers, in GaAs.

Where ΔS_{cv} denotes the combined effect of one e_c and one h_v, it is sometimes possible to separate the two effects by observing the isotope shifts of zero-phonon optical transitions due to impurities [28, 29]. One may deduce from the isotope shift the local softening effect on the phonon modes about the impurity when one electron or one hole localises into a state about the impurity. Heine and Henry studied this problem [28] for several semiconductors, particularly GaP and ZnO. They concluded that, at least for these cases, a hole is almost four times as effective in softening phonons as is an electron. That is

$$S_h = (3.6 \pm 1)S_e \tag{9}$$

with

$$\Delta S_{cv} = S_e + S_h. \tag{10}$$

As suggested by Heine and Henry, we assume that this distribution of the weight of the two contributions is essentially the same for all tetrahedrally bonded semiconductors.

Let us now define the band offset problem in thermodynamic terms. (Care is required not to make a sign error.) We have the two semiconductors, A and B, and consider the reaction that transfers an e_c from B to A at the cost of free energy $\Delta E_c(A/B)$,

$$A{:}0,\ B{:}e_c^- \Rightarrow A{:}e_c^-,\ B{:}0 \qquad \Delta E_c(A/B) = E_c(A) - E_c(B). \tag{11}$$

We also have the reaction that transfers a h_v from B to A at the cost of free energy $-\Delta E_v(A/B)$ (note the effect of the difference in the sign of the charge of the h_v and the e_v),

$$A{:}0,\ B{:}h_v^+ \Rightarrow A{:}h_v^+,\ B{:}0 \qquad -\Delta E_v(A/B) = E_v(A) - E_v(B). \tag{12}$$

The vacuum level, or any other reference state, R, may be introduced by breaking these reactions into two parts, i.e.

$$B{:}e_c^-,\ R{:}0 \Rightarrow B{:}0,\ R{:}e^- \tag{13}$$

followed by

$$R{:}e^-,\ A{:}0 \Rightarrow R{:}0,\ A{:}e_c^- \tag{14}$$

and correspondingly for the hole reaction. Adding reactions (11) and (12), we have

$$A{:}0,\ B{:}e_c^- + h_v^+ \Rightarrow A{:}e_c^- + h_v^+,\ B{:}0 \qquad \Delta E_c(A/B) - \Delta E_v(A/B) \tag{15}$$

which immediately implies

$$\Delta E_{cv}(A) - \Delta E_{cv}(B) = \Delta E_c(A/B) - \Delta E_v(A/B) \tag{16}$$

as claimed above at equation (5). When we differentiate equation (16) with respect to T, we have

$$\mathrm{d}\,\Delta E_{cv}(A)/\mathrm{d}T - \mathrm{d}\Delta E_{cv}(B)/\mathrm{d}T = \mathrm{d}\,\Delta E_c(A/B)/\mathrm{d}T - \mathrm{d}\,\Delta E_v(A/B)/\mathrm{d}T. \tag{17}$$

Now we recall the basic thermodynamic identity that the entropy is the negative of the derivative of the corresponding free energy (or chemical potential) with respect to T, so

$$S_c(A) = -\mathrm{d}E_c(A)/\mathrm{d}T \qquad S_v(A) = -\mathrm{d}E_v(A)/\mathrm{d}T \tag{18}$$

and, of course, the same for B while for the band gaps themselves

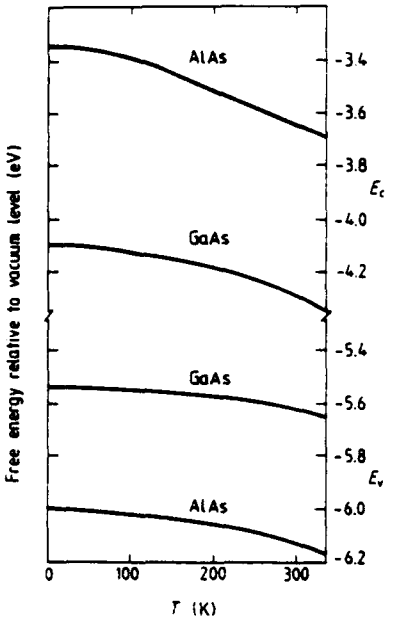


Figure 3. Variation with T of the valence and conduction band edge free energies, E_v and E_c, for GaAs and AlAs.

$$\Delta S_{cv}(A) = -dE_{cv}(A)/dT. \tag{19}$$

Finally, when we compare equation (10) with (18), we have the interesting result that

$$S_e + S_h = S_c - S_v \tag{20}$$

or

$$S_c = \Delta S_{cv} + S_h. \tag{21}$$

This merely states the fact that, in order for the band gap to decrease with rising T, the absolute free energy E_c must fall faster than the absolute free energy E_v by ΔS_{cv}. This is illustrated in figure 3 for the cases of GaAs and AlAs.

We now introduce the physical assumption of equation (9), that the fraction of ΔS_{cv} due to h_v is a constant 77% (± 5) for all tetrahedrally bonded semiconductors, and we equate this with S_v:

$$S_v = S_h \tag{22}$$

so we have

$$d\ \Delta E_v(A/B)/dT = 0.77\ (\pm 0.05)(d\ \Delta E_{cv}(A)/dT - d\ \Delta E_{cv}(B)/dT). \tag{23}$$

While equation (22) is attractive on intuitive physical grounds, it is not clear whether it is rigorously exact. The biggest problem is that it assumes that the effect on the lattice modes of a localised hole or electron, as observed in the isotope shift experiments [28], is the same as that of a delocalised h_v or e_c in the band edge density of state distribution. Van Vechten and Thurmond have argued that this should be so to a good approximation [22, 29] and correlated it to the proposition that for any tetrahedral semiconductor the temperature dependence of the various direct band gaps in the optical spectrum (the fundamental gap and all higher gaps) should be the same. (This implies that a delocalised

state at any point in the Brillouin zone has the same effect on the lattice, so the effect of a localised state, having components from many points in the zone, would also be the same.) What experiments have been done to test this hypothesis [30] (on Si) support it. Aside from this problem of a possible distinction between localised and delocalised carriers, one can question the consequences of the possible alternative assumptions that $S_v < S_h$ or $S_v > S_h$. If $S_v < S_h$, one would have to conclude that the addition of antibonding charge, i.e. the e_c, someshow stiffens the lattice relative to the degree indicated by the isotope shift experiments—so the entropy of the valence band edge distribution could be reduced. This seems most unlikely. If $S_v > S_h$, then there must be some contribution to the entropy of both the valence and conduction band density of states beyond that due to the lattice modes. (Recall the quantitative description of the variation of the mode frequencies [26] to the melting points of Si and Ge.) No source of such a contribution is apparent.

Let us now note the consequence of equations (21), (22) and (23) for ΔE_c:

$$\mathrm{d}\,\Delta E_c(\mathrm{A/B})/\mathrm{d}T = 1.77\,(\pm 0.05)(\mathrm{d}\,\Delta E_{cv}(\mathrm{A})/\mathrm{d}T - \mathrm{d}\,\Delta E_{cv}(\mathrm{B})/\mathrm{d}T). \tag{24}$$

Furthermore, because ΔH_V and ΔH_c are connencted by thermodynamic identities [21] to ΔS and ΔE_c and ΔE_v:

$$\mathrm{d}\,\Delta H_v(\mathrm{A/B})/\mathrm{d}T = 0.77\,(\pm 0.05)(\mathrm{d}\,\Delta H_{cv}(\mathrm{A})/\mathrm{d}T - \mathrm{d}\,\Delta H_{cv}(\mathrm{B})/\mathrm{d}T) \tag{25}$$

and

$$\mathrm{d}\,\Delta H_c(\mathrm{A/B})/\mathrm{d}T = 1.77\,(\pm 0.05)(\mathrm{d}\,\Delta H_{cv}(\mathrm{A})/\mathrm{d}T - \mathrm{d}\,\Delta H_{cv}(\mathrm{B})/\mathrm{d}T). \tag{26}$$

To carry this discussion any further, we must consider specific cases. We find data adquate for a discussion available for two cases—those of GaAs–AlAs junctions and of HgTe–CdTe functions, and that these generally support the foregoing simple theory.

3. The case of AlAs–GaAs

In the AlAs–GaAs heterojunction, E_v for GaAs is higher [31] than for AlAs at $T = 0$. It was found in [32] that $\mathrm{d}\,\Delta E_{cv}/\mathrm{d}T$ is a linear function of the mole fraction of Al in the $Al_xGa_{1-x}As$ alloy system for $x < 0.50$. We use their extrapolation to AlAs, which concluded that $\mathrm{d}\,\Delta E_{cv}/\mathrm{d}T$ for AlAs is 1.6 ± 0.2 times that for GaAs. Consequently, E_v(GaAs) will decrease with rising T more slowly than will E_v(AlAs), and ΔE_v(GaAs/AlAs) will become larger. (See figure 3.) However, ΔH_v(GaAs/AlAs), which equals ΔE_v(GaAs/AlAs) at $T = 0$ K, will decrease with rising T because H_v(GaAs) will increase more slowly than H_v(AlAs).

Consider now what empirical information relevant to the simple theory of equations (23) and (24) is available. The theory implies that for most cases $\mathrm{d}E_V(\mathrm{A/B})/\mathrm{d}T$ and $\mathrm{d}E_c(\mathrm{A/B})/\mathrm{d}T$ are both small because the $\mathrm{d}\Delta E_{cv}(\mathrm{A})/\mathrm{d}T \simeq \mathrm{d}\Delta E_{cv}(\mathrm{B})/\mathrm{d}T$ for most pairs of semiconductors A and B. Furthermore, the total variation of ΔE_{cv}(A) over the range of accurate experiments, generally from $T = 0$ K to 300 K, is not much larger than the experimental uncertainty in ΔE_v(A/B) or ΔE_c(A/B). (For GaAs, $\Delta E_{cv}(T = 0\ \mathrm{K}) - \Delta E_{cv}(T = 295\ \mathrm{K}) = 94$ meV, while $\Delta H_{cv}(T = 295\ \mathrm{K}) - \Delta H_{cv}(T = 0\ \mathrm{K}) = 39$ meV.) This is certainly consistent with the fact that the authors have not been able to find any explicit discussion of the temperature dependence of band offsets in the previous

literature. However, one can look at discrepancies in determinations made at different values of T.

The case of AlAs–GaAs heterojunctions has probably been studied more carefully and with better prepared samples than any other. For the cases $Al_xGa_{1-x}As$ alloys on GaAs, Wolford *et al* determined [31] that $\Delta E_v(T = 8\,K) = 110 \pm 8$ meV for GaAs–$Al_{0.28}Ga_{0.72}As$ and $\Delta E_v(T = 8\,K) = 320 \pm 10$ meV for GaAs–$Al_{0.70}Ga_{0.30}As$. They used optical methods, it is shown in [22] that such optical experiments do in fact measure chemical potentials and thus, as Wolford *et al* imply, do determine ΔE_vs or ΔE_{cv}s rather than ΔH_vs or ΔH_{cv}s. Another careful determination of similarly well prepared samples was made [33] for higher T by Batey and Wright. They studied the thermionic emission of h_vs across the junction in p-type material as a function of alloy composition and of T for the interval 79.6 K to 294.4K. They concluded that

$$\Delta E_v(GaAs/Al_xGa_{1-x}As) = 0.55x\ eV\ (\pm 20 - 40\ meV) \tag{27}$$

where x is the mole fraction of Al in the alloy. This implies that $\Delta E_v = 154 \pm 30$ meV for GaAs–$Al_{0.28}Ga_{0.72}As$ in contrast to Wolford and co-workers' 110 ± 8 meV, and 385 ± 30 meV for GaAs–$Al_{0.70}Ga_{0.30}As$ in contrast to 320 ± 10 meV.

Batey and Wright obtained equation (27) by fitting their data for the thermionic emission current, J_s, with a Richardson equation

$$J_s = A^*T^2 \exp(-\phi/kT) \tag{28}$$

where A^* is the Richardson constant and s the activation barrier for holes. The limit of ϕ as bias voltage goes to zero can be simply related to ΔE_v, if, as Batey and Wright did, one takes account of the temperature variation of the Fermi level. However, in their analysis of Arrhenius plots of ϕ versus $1/T$, ΔE_v was treated as if it were independent of T. A reanalysis of the raw data including the expected variation of ΔE_v (and thus of ϕ) with T is required.

Although the issue may now be unresolvable, we can discuss the implications of the temperature dependence of ΔE_v regarding a reconciliation of these two experiments. First let us note that, since ϕ is a free energy barrier, as is ΔE_v, Batey and Wright should have measured a larger value than Wolford *et al* simply because, as noted above, the facts that $d\Delta E_{cv}/dT$ is greater for AlAs than for GaAs and that $E_v(GaAs) > E_v(AlAs)$ imply that $\Delta E_v(GaAs/AlAs)$ increases with T. For the case of the 70% AlAs alloy, we have

$$\Delta E_{cv}(T = 0\,K) - \Delta E_{cv}(T = 295\,K) = 94(1 + 0.7\,(0.6 \pm 0.2))\ meV = 134 \pm 13\ meV. \tag{29}$$

Thus, ΔE_{cv} decreases by 40 ± 13 meV more in the alloy than in pure GaAs and, by our argument, $77 \pm 5\%$ of this difference occurs at the valence band edge. Thus, we predict that ΔE_v should be as much as 31 ± 10 meV greater at 295 K that at 0 K. Depending on the details of Batey and Wright's analysis of their data, then one sees that the expected variation of ΔE_v with T accounts for 31 ± 10 meV of the 65 ± 40 meV discrepancy between that determination and the optical experiment [31] at a constant $T = 8$ K.

A final resolution of this descrepancy will require a reanalysis of the raw data of the thermionic emission study, which is not available to the present authors. However, it would appear that the consequences of the variation of ΔE_v with t have the approximate magnitude and arguably the correct sign to reconcile these two fine experiments [31, 33].

4. The case of HgTe–CdTe

Let us turn to a case where the effect of T should be expected to be larger and to be more clearly recognised. Such a case, where the offsets have also received much experimental attention, is the HgTe–CdTe heterojunction [19, 20, 34, 35]. There is a large literature on this heterojunction due to the effort to develop infra-red devices based upon it.

The reader should note that, because HgTe is a semimetal, the antibonding-s Γ_6 level lies below the bonding-p Γ_8 level. However, by convention, one continues to label Γ_6 as 'E_c' and Γ_8 as 'E_v'.

Kowalczyk *et al* concluded from a room temperature optical experiment [34] that $\Delta E_v(\text{HgTe/CdTe}) = 0.35 \pm 0.06$ eV, i.e. that the HgTe valence band edge is higher than that of CdTe by this amount. This conclusion has been controversial because while others have repeated Kowalczyk and co-workers' experiment [36], several other determinations of the band offset have implied values an order of magnitude lower (see the discussion in [19]).

First we note that the experiments giving the smaller values for ΔE_v were all done at low temperatures (around 4 K). Some of these estimated ΔE_v from the optical properties of superlattice structures. The methods used for those inferences are somewhat controversial. We choose to avoid that controversy by instead basing our discussion on estimates of ΔE_v derived entirely from electronic transport measurements [37]. Chow *et al* deduced [35] $\Delta E_v(\text{HgTe/CdTe}) < 0.10$ eV at $T = 4$ K using the same structure that demonstrated [20] $\Delta E_v(\text{HgTe/CdTe}) = 0.35 \pm 0.06$ eV at $T = 300$ K, consistent with [34]. In particular, they demonstrated negative differential resistance in $Hg_{0.78}Cd_{0.22}Te$/CdTe junction device [35] at $T = 4$ K and argued forcefully that this observation requires $\Delta E_v(\text{HgTe/CdTe}) < 0.10$ eV.

We now consider whether the discrepancy in $\Delta E_v(\text{HgTe/CdTe})$ can be ascribed to the expected temperature variation of the two E_vs. We first note that HgTe has the property $\mathrm{d}\,\Delta E_{cv}/\mathrm{d}T > 0$ in sharp contrast to the $\mathrm{d}\,\Delta E_{cv}/\mathrm{d}T < 0$ behaviour of CdTe, GaAs, AlAs, Si, and most other semiconductors. As shown by Heine and Van Vechten, this is because [23] the top of the valence band in HgTe consists of states with antibonding-s character, rather than bonding-p character as with the other semiconductors; the states at the bottom of the conduction band have bonding-p character. Thus, the effect of thermally excited e_cs and h_vs is to stiffen, rather than to soften, the phonon modes as $\Delta S_{cv} < 0$. This means that for HgTe, and its alloys, E_v rises with increasing T while for CdTe, E_v falls. (See figure 4.) It then follows from equation (23) that if $E_v(\text{HgTe}) > E_v(\text{CdTe})$ at $T = 0$, then $\Delta E_v(\text{HgTe/CdTe})$ will increase rapidly with T for $T > 0$. (If $E_v(\text{HgTe}) < E_v(\text{CdTe})$ at $T = 0$, then with increasing T they would cross and $E_v(\text{HgTe})$ will rapidly rise above $E_v(\text{CdTe})$.) The variation is particularly rapid because both terms in equation (23) are positive; there is no cancellation between them. Indeed, we expect

$$\begin{aligned}\Delta E_v(\text{HgTe/CdTe}, T = 300\text{ K}) - \Delta E_v(\text{HgTe/CdTe}, T = 0\text{ K}) \\ = 0.77[(\Delta E_{cv}(\text{HgTe}, T = 300\text{ K}) \\ - \Delta E_{cv}(\text{HgTe}, T = 0\text{ K})) + (\Delta E_{cv}(\text{CdTe}, T = 0\text{K}) \\ - \Delta E_{cv}(\text{CdTe}, T = 300\text{ K}))]. \qquad (30)\end{aligned}$$

To be quantitative, $\Delta E_{cv}(\text{HgTe})$ increases [38] by about 0.160 eV between 0 and 300 K

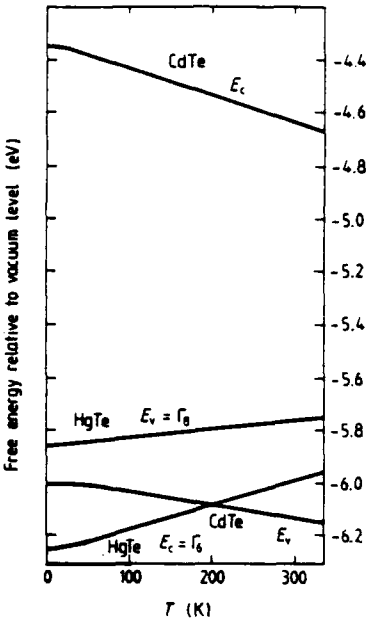


Figure 4. Variation with T of the valence and conduction band edge free energies for HgTe–CdTe. In this figure we have assumed ΔE_v(HgTe/CdTe) = 50 meV at $T = 0$ K and the variation with T of ΔE_{cv} for both HgTe and CdTe to be as reported in [38]. The HgTe bands are labelled according to the prevailing 'negative band gap' convention also used in [34–39]. Thus, the antibonding-s Γ_6 level, which corresponds to the bottom of the conduction band in CdTe (and in GaAs), is labelled 'E_c' even though it lies below the bonding-p Γ_8 level, which corresponds to the top of the valence band in CdTe (and in GaAs, AlAs etc), and which is labelled 'E_v'. The ΔE_v values quoted are indeed the differences of the Γ_8 levels across the junction.

while ΔE_{cv}(CdTe) decreases [38] about 0.161 eV. Thus, the expected variation of ΔE_v is

$$\Delta E_v(\text{HgTe/CdTe}, T = 300\text{ K}) - \Delta E_v(\text{HgTe/CdTe}, T = 0\text{ K}) = 0.247\text{ eV}. \tag{31}$$

To this can be added the effect of the alloy variation of E_v; Chow *et al* set their lower limit E_v on the basis of a 78% Hg alloy rather than for pure HgTe. Approximating the variation of the E_v as linear, we estimate that if $\Delta E_v(T = 0) = 0.10$ eV for the 78% alloy, as Chow *et al* allow, then it is 0.13 eV for pure HgTe. (This should be a small overestimate due to the alloy disorder induced bowing of the band edges [22, 38].) We would then calculate

$$\Delta E_v(\text{HgTe/CdTe}, T = 300\text{ K}) = 0.13 + 0.25 = 0.38\text{ eV} \tag{32}$$

which is even larger than estimated by Kowalczyk *et al* but within their error limit, ΔE_v(HgTe/CdTe, $T = 300$ K) = 0.35 ± 0.06 eV. Obviously, the value calculated at equation (32) can be reduced by allowing for the alloy bowing (an effect estimated to be roughly 0.01 eV) and by assuming $\Delta E_v(T = 0)$ to be less than the maximum allowed by Chow *et al*. In addition, the variation of ΔE_{cv}(CdTe) may be less than reported in [38] and assumed here. Zanio reported [39] the variation to be only 0.10 eV from 0 to 300 K, which would reduce the estimate of equation (32) to 0.33 eV.

5. Summary

We conclude that this variation with T is the likely explanation for the discrepancy between the experimental determinations of ΔE_v(HgTe/CdTe) by Chow *et al* and by Kowalczyk *et al*; both results can be quite correct. Furthermore, the very simple theory proposed here is consistent with all data available to us.

We have presented the thermodynamic perspective on the temperature dependence of the band offsets in heterojunctions. We have argued that the magnitude of the variation is directly proportional to the difference in the band gap temperature variation. The HgTe–CdTe and AlAs–GaAs systems were specifically discussed, but extensions to other materials combinations and to the effect of pressure on the band offsets follow in similar fashion.

Acknowledgments

We are grateful to D J Wolford for very useful discussions of the AlAs–GaAs case. The work at Oregon State University was supported in part by the Air Force of Scientific Research under Contract No AFOSR 86-0309.

References

[1] Kromer H 1957 *Proc. Inst. Radio Eng.* **45** 1535
[2] Rupprecht H S, Woodall J M and Pettit G D 1967 *Appl. Phys. Lett.* **11** 81
[3] Alferov Zh, Andreev V M, Korol'kov V I, Portnoi E L and Tret'yakov D N 1969 *Fiz. Tekh. Poluprov.* **2** 1016 (Engl. Transl. *Sov. Phys.–Semicond.* **2** 843)
[4] Panish M B and Sumski S 1969 *J. Phys. Chem. Solids* **30** 129
[5] Esaki L and Tsu R 1970 *IBM J. Res. Dev.* **14** 61
[6] Dingle R, Stromer H L, Gossard A C and Wiegmann W 1978 *Appl. Phys. Lett.* **33** 665
[7] Kromer H 1984 *Molecular Beam Epitaxy and Heterostructures* ed L L Chang and K Ploog (The Hague: Nijhoff)
[8] Arthur J R (ed) 1985 *Proc. 3rd Int. Molecular Beam Epitaxy Conf.*; *J. Vac. Sci. Technol.* B **3** 1
[9] Casselman T N 1985 *Proc. 1985 US Workshop Physics and Chemistry of Mercury Cadmium Telluride. J. Vac. Sci. Technol.* A **3** 47
[10] Dingle R (ed) 1987 *Applications of Multiquantum Wells, Selective Doping, and Superlattices* (New York: Academic)
[11] Herring C and Nichols M H 1949 *Rev. Mod. Phys.* **21** 185
[12] Anderson R L 1962 *Solid State Electron.* **5** 341
[13] Van Vechten J A 1969 *Phys Rev.* **182** 891
[14] Tersoff J 1984 *Phys. Rev.* B **30** 4874
[15] Van Vechten J A 1985 *J. Vac. Sci. Technol.* B **3** 1240
[16] Cardona M and Christensen N E 1987 *Phys. Rev.* B **35** 6182
[17] Van de Walle C G and Martin R M 1987 *Phys. Rev.* B **35** 8154
[18] Bylander D M and Kleinman L 1987 *Phys. Rev. Lett.* **18** 2091
[19] Faurie J-P, Hsu C and Duc T-M 1987 *J. Vac. Sci. Technol.* A **5** 3074
[20] Chow D H, McCaldin J O, Bonnefoi A R, McGill T C, Sou I K and Faurie J-P 1987 *Appl. Phys. Lett.* **51** 2230
[21] Thurmond C D 1975 *J. Electrochem. Soc.* **122** 1133
[22] Van Vechten J A 1980 *Handbook on Semiconductors, Vol. 3, Materials, Properties and Preparations* ed S P Keller (Amsterdam: North-Holland) ch 1
[23] V Heine and Van Vechten J A 1976 *Phys. Rev.* B **13** 1622
[24] Martin R M 1969 *Phys. Rev.* **186** 871
[25] Weber W 1974 *Phys. Rev. Lett.* **33** 371

[26] Van Vechten J A 1985 *Phys. Rev.* B **31** 2508
[27] Vina L, Hochst H and Cardona M 1985 *Phys. Rev.* B **31** 958
[28] Heine V and Henry C H 1975 *Phys. Rev.* B **11** 3795
[29] Van Vechten J A and Thurmond C D 1976 *Phys. Rev.* B **14** 3539
[30] Van Vechten J A 1976 *Phys. Rev.* B **13** 946; 1976 *Phys. Rev.* B **14** 1781
[31] Wolford D J, Keuch T F and Bradley J A 1986 *J. Vac. Sci. Technol.* B **4** 1043
[32] Vorobkalo F M, Glinchuk K D and Kovalenko V F 1975 *Fiz. Tekh. Poluprov.* **9** 998 (Engl. Transl. 1975 *Sov. Phys. Semicond.* **9** 656)
[33] Batey J and Wright S L 1986 *J. Appl. Phys.* **59** 200
[34] Kowalczyk S P, Cheung J T, Kraut E A and Grant R W 1986 *Phys. Rev. Lett.* **56** 1605
[35] Chow D H, McGill T C, Sou I K, Faurie J P and Nieh C W 1988 *Appl. Phys. Lett.* **52** 54
[36] Shih C K and Spicer W E 1987 *Phys. Rev. Lett.* **58** 2594
[37] Malloy K J and Van Vechten J A 1989 *Appl. Phys. Lett.* **54** 937
[38] Hansen G L, Schmidt J L and Casselman 1982 *J. Appl. Phys.* **53** 7099
[39] Zanio K 1978 *Semiconductors and Semimetals* ed R K Willardson and A C Beer (New York: Academic) p 99

Thermal Expansion Contributions to Band Gap and Band Offset Temperature Dependences.

Kevin J. Malloy
The Center for High Technology Materials
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131-6081

and

James A. Van Vechten
Center for Advanced Materials Research
Department of Electrical and Computer Engineering
Oregon State University
Corvallis, OR 97331-3211

ABSTRACT

Using a thermodynamic approach, we examine the role thermal expansion plays in determining the temperature dependence of band gaps and band offsets. Because of the large variation in the standard volume of electrons (or, equivalently, variation in conduction band deformation potential) with symmetry of the conduction band, most of the band gap temperature variation can be ascribed to the valence band for the X gap, while the variation is more equitably divided between the conduction and valence band for the Γ band gap. The particulars of the AlAs/GaAs system are examined in detail, where we find evidence that X electrons have a larger electron-phonon entropy than Γ electrons, and that the lowest conduction band offset exhibits considerable temperature variation.

Many device concepts make use of transitions between energy levels in semiconductor structures. Given that a real device will usually have to operate over some temperature range, understanding of the temperature dependence of the energy levels is key to furthering the technology of such a device. While the temperature dependences of band gaps are well known, only recently have the temperature dependences of the band offsets that play a crucial role in quantum well and other heterostructure devices been investigated[1,2]. Here we elaborate on our previous exposition by including a discussion of the role thermal expansion effects play in the temperature dependences of band offsets.

Our approach has been to apply the insights derived from the bond charge model of the solid[3,4] to a thermodynamic analysis of the problem. We begin by introducing the thermodynamics of energy gaps as first discussed by Brooks (Ref. 5), Thurmond (Ref. 6) and others (Ref. 7). The band gap ΔE_{cv}, is defined as the standard free energy of formation for an electron-hole pair;

$$0 \rightarrow e_{cb} + h_{vb}; \qquad \Delta E_{cv} \qquad (1),$$

where e_{cb} refers to an electron in the conduction band and h_{vb} refers to a hole in the valence band.

In the conventional, single electron band diagram picture, ΔE_{cv} is defined as the difference between the conduction band edge E_c, and the valence band edge E_v. This view is reconciled with Eq. (1) if we remember thermodynamic values are defined for a statistical ensemble, while most one electron energy band calculations are defined for T = 0K (Ref. 7). Also, a profound difference exists between terms such as the "total (free) energy" used in association with numerical calculations of a system and the "standard free energy change" of a reaction such as Eq. (1) above.

To investigate the temperature dependence of the band edges (and ultimately the band offset), we divide Eq. (1) into two parts:

$$h_{vac} \rightarrow h_{vb}, \qquad -E_v \qquad (2), \text{ and}$$

$$e_{vac} \rightarrow e_{cb}, \qquad E_c \qquad (3).$$

Now h_{vac} and e_{vac} refer to holes and electrons in some reference state, typically as we'll describe below, the vacuum level.

We further convert Eq. (2) into Eq. (2a), describing the hole transition from the electron's perspective as

$$e_{vb} \rightarrow e_{vac}, \qquad -E_v \qquad (2a),$$

to help avoid making any sign errors or misinterpretations. Many such opportunities arise when trying to interpret thermodynamic relationships in terms of a one-electron energy level diagram or when trying to understand

electron energy levels in a thermodynamic context. We also have left out the "Δ" in front of the free energy changes for Eqs. (2-3) partially to avoid confusion with band offsets and partially to anticipate a connection with the individual band edges. Still, E_c and E_v are not single electron energy states, but are instead the free energy of reaction required to place an electron or hole in their standard state, which as Ref. 6 describes, is the conduction band edge or valence band edge, respectively.

Therefore, the reassuring sum of Eq. (2) and (3) is

$$\Delta E_{cv} = E_c - E_v, \qquad (4),$$

To assign electron energy levels, some reference level must be postulated that can interact with the semiconductor so as to leave the electron in the desired state. Traditionally the vacuum level, defined as the energy of an electron an infinite distance away from the sample, serves as this reference. Experimentally, the vacuum reference level is hopelessly inaccessible; surface dipole layers complicate the experimental measurement of this level and image and other electrostatic forces associated with the rearrangement of charge also obscure experiments[8]. Therefore, the vacuum reference serves as an ideal rather than a real level with the concomitant implication that the band edges, band gaps and band offsets are idealized, intrinsic bulk properties independent of surfaces and interfaces. It should be noted that evidence is mounting about the strong role of the surface and interface in such properties as band offsets. However, recent studies[9-11] have also suggested that certain deep impurities in semiconductors can serve suitably well as this reference level.

Figure 1 summarizes the ideas used in deriving Eqs. (1-4) for a conventional semiconductor such as GaAs or AlAs. In Fig. 1, we qualitatively depict the temperature dependence of the standard free energy of formation for the indicated reactions based on the vacuum level as the reference level. We note that Eq. (2a) is not a spontaneous reaction and therefore has a positive standard free energy, while the reverse is true for Eq. (3). Thus, E_c and E_v are negative quantities with respect to the vacuum level. In addition, ΔE_{cv}, E_v and E_c will all follow Gibb's equation;

$$\Delta E_{cv} = \Delta H_{cv} - T\Delta S_{cv} \qquad (5).$$

Here, ΔH_{cv} is the standard enthalpy and ΔS_{cv} the standard entropy of formation for electron-hole pairs. Again, we need to make careful distinctions between assessments of standard internal energy, free enthalpy, or free entropy and the T=0K solutions to Schrödinger's equation when discussing these issues.

Standard Entropy and Standard Volume of Formation

Once the connection is made between the single electron eigenvalues and the thermodynamic quantities actually measured by experiments, we can begin to

understand their temperature and pressure dependences. The temperature dependence of ΔE_{cv} is defined by the relationship

$$\left.\frac{d\Delta E_{cv}}{dT}\right|_P = -\Delta S_{cv}. \qquad (6).$$

The corresponding relationship for the pressure dependence of ΔE_{cv} is given by

$$\left.\frac{d\Delta E_{cv}}{dP}\right|_T = \Delta V_{cv}. \qquad (7).$$

where ΔV_{cv}.is the standard volume change for creation of an electron-hole pair. Before we begin our discussion of band offsets and the individual contributions of electrons and holes, it is useful to further review the thermodynamics of band gaps.

Separation of Electron-Phonon and Thermal Expansion Components

Under the constant pressure conditions of most experiments, the standard entropy is given by

$$-\left.\frac{d\Delta E_{cv}}{dT}\right|_P = \Delta S_{cv} = -\left.\frac{\partial \Delta E_{cv}}{\partial T}\right|_V - \left.\frac{\partial V}{\partial T}\right|_P \left.\frac{\partial \Delta E_{cv}}{\partial V}\right|_T \qquad (8),$$

$$= -\left.\frac{\partial \Delta E_{cv}}{\partial T}\right|_V - \left.\frac{\partial V}{\partial T}\right|_P \left.\frac{\partial P}{\partial V}\right|_T \left.\frac{\partial \Delta E_{cv}}{\partial P}\right|_T \qquad (8a),$$

where $\frac{1}{V}\left.\frac{\partial V}{\partial T}\right|_P = 3\alpha_p$, with α_p the linear expansion coefficient at constant pressure, and $V\left.\frac{\partial P}{\partial V}\right|_T = -\beta_T$, the negative of the isothermal bulk modulus or the inverse negative of the isothermal compressibility (the ratio of the adiabatic bulk modulus, β_s, to the isothermal bulk modulus, β_s/β_T is given by C_p/C_v exactly or by $(1+T\gamma 3\alpha_p)$ for an isotropic Debye solid, where γ is the Gruineissen constant and C_p and C_v are the heat capacities at constant pressure and volume, respectively. Both factors both near 1 for semiconductors like GaAs). The two terms on the right hand side of Eq. (8) are somewhat imprecisely referred to as the electron-phonon entropy and the thermal expansion entropy respectively.

Previously[1,2], we investigated the first term on the right of Eq. (8), the electron phonon contribution to the entropy. We pointed out that earlier analyses[12] of experiments done at liquid Helium temperatures, show that if we write

$$\Delta S_{cv}^{(elec\text{-}phon)} = -\left.\frac{\partial \Delta E_{cv}}{\partial T}\right|_V = S_e^{(elec\text{-}phon)} + S_h^{(elec\text{-}phon)} \qquad (9),$$

where S_e is the standard free entropy of an electron and S_h the standard free entropy of a hole, then for GaP, the relationship

$$S_h^{(elec\text{-}phon)} = 3.6(\pm 1) S_e^{(elec\text{-}phon)} \qquad (10)$$

holds. A larger hole than electron contribution to $\Delta S_{cv}^{(elec\text{-}phon)}$ is not only consistent with recent theoretical calculations on Si and Ge[13] but is understood in terms of the bond charge model as arising from the sensitivity of TA phonons to the charge present in the bond[14] and from the lack of coupling between TA phonons and spherically symmetric electron states. Thus, removing an electron from a bonding state softens the lattice much more than placing an electron in an antibonding state.

Using Eq. (4), we have

$$S_e + S_h = -\frac{d(E_c - E_v)}{dT} = -\frac{dE_c}{dT} + \frac{dE_v}{dT} \qquad (11)$$

or, combining with Eq. (10),

$$S_h^{(elec\text{-}phon)} = -0.77(\pm 0.05)\left.\frac{\partial \Delta E_{cv}}{\partial T}\right|_V$$

$$S_e^{(elec\text{-}phon)} = -0.23(\pm 0.05)\left.\frac{\partial \Delta E_{cv}}{\partial T}\right|_V \qquad (12).$$

If we now consider what happens at band offsets, we can follow Ref. 2 and write the reaction for the transfer of an e-h pair from semiconductor B to semiconductor A as:

$$A{:}0,\ B{:}(e_c^- + h_v^+) \rightarrow A{:}(e_c^- + h_v^+),\ B{:}0 \qquad \Delta E_c(A/B) + \Delta E_v(A/B) \qquad (13),$$

with the reaction described by Eq. (13) broken down into two half-reactions at the band edges:

$$A{:}0,\ B{:}h_{vb} \rightarrow A{:}h_{vb},\ B{:}0 \qquad \Delta E_v(A/B) = -(E_v(A) - E_v(B)) \qquad (14), \text{ and}$$

$$A{:}0,\ B{:}e_{cb} \rightarrow A{:}e_{cb},\ B{:}0 \qquad \Delta E_c(A/B) = E_c(A) - E_c(B) \qquad (15).$$

In order to extend our discussion to band offsets, we must separate out the thermal expansion contribution of each band edge. When discussing the variation of band edges with applied stress, it is common to use the deformation potential concept[16]. That is equally applicable here since:

$$\Delta S_{cv}(\text{Therm Expan}) = -\left.\frac{\partial V}{\partial T}\right|_P \left.\frac{\partial P}{\partial V}\right|_T \left.\frac{\partial \Delta E_{cv}}{\partial P}\right|_T$$

$$= 3\alpha_P\beta_T\Delta V_{cv} = -3\alpha_P a_{cv} \qquad (20)$$

where a_{cv} is the band gap deformation potential. We list the following useful relations to connect the more familiar deformation potential with our concept of standard volume:

$$a_{cv} = a_c - a_v = -\beta_T\Delta V_{cv} = -\beta_T(V_e + V_h) \qquad (21),$$

$$\frac{d\Delta E_{cv}}{dP} = \Delta V_{cv} = V_e + V_h = \frac{dE_c}{dP} - \frac{dE_v}{dP} \qquad (22), \text{ and}$$

$$-\frac{a_c}{\beta_T} = V_e = \frac{dE_c}{dP},$$

$$\frac{a_v}{\beta_T} = V_h = -\frac{dE_v}{dP} \qquad (23).$$

Our observation shall be that a_c, the conduction band deformation potential, varies strongly with the symmetry of the conduction band involved. Therefore, the standard volume of creation of electrons also will depend on the conduction band involved and the thermal expansion contribution to the entropy of conduction band states also will vary accordingly.

First, let us make the point that these volumes are readily observable effects. Several experimental results that involve placing free carriers in the semiconductor show measurable volume changes to the crystal as a whole consistent with the appropriate deformation potential/standard volume. For instance, generation of electron-hole pairs through illumination[17,18], doping of semiconductors n- or p-type[19,20] and the transitions involving defects[21,22] all result in measurable expansion or contraction of the lattice. Thus we may (somewhat imprecisely) speak of ΔV_{cv} as the volume of an electron-hole pair and V_e and V_h as the volume of an electron and hole respectively.

To summarize by combining the results of this section with the results of the previous section, for a given semiconductor, we may write:

$$\Delta S_{cv}(\text{total}) = \Delta S_{cv}(\text{elec-phon}) + 3\alpha_P\beta_T\Delta V_{cv},$$
$$S_e(\text{total}) = S_e(\text{elec-phon}) + 3\alpha_P\beta_T V_e, \text{ and}$$

$$S_h^{(total)} = S_h^{(elec\text{-}phon)} + 3\alpha_P\beta_T V_h \quad (24)$$

for the band gap and

$$\Delta S_c^{(A/B)} = S_e^A(elec\text{-}phon) - S_e^B(elec\text{-}phon) + 3\alpha_P^A\beta_T^A V_e^A - 3\alpha_P^B\beta_T^B V_e^B$$
$$\Delta S_v^{(A/B)} = S_h^A(elec\text{-}phon) - S_h^B(elec\text{-}phon) + 3\alpha_P^A\beta_T^A V_h^A - 3\alpha_P^B\beta_T^B V_h^B \quad (25)$$

for the band offsets.

According to the "Empirical Rule,"[15] the following relationships for the various conduction band gaps generally hold true for III-V (and other) semiconductors:

$$\Delta V_{cv}(\Gamma) \approx 12\text{-}15\frac{meV}{kbar} = 19\text{-}24\ \text{Å}^3$$
$$\Delta V_{cv}(X) \approx -1.5\frac{meV}{kbar} = -2.4\ \text{Å}^3$$
$$\Delta V_{cv}(L) \approx 5\frac{meV}{kbar} = 8\ \text{Å}^3.$$

The band gap deformation potentials then vary according to Eq. (21). Enough data exists on the band edge deformation potentials to suggest a_c^Γ is large and negative, a_c^X is small and a_c^L is moderately negative. The data on a_v clearly suggest it is small, but are mixed on the sign. We choose to believe a_v has a small negative sign (and therefore $V_h < 0$) based on three different arguments: If PV work was done on a semiconductor at 0K, one would expect the valence electrons to have some compressibility and to therefore store some energy. Therefore, a transition involving a valence band would decrease in absolute free energy as pressure is applied to the semiconductor. This is equivalent to saying $dE_v/dP > 0$ or that a hole has negative volume. The second argument states that the zone center valence band edge states are p-type bonding states. Upon removal of charge from such a state, the atoms will move closer since the sp^3 states comprising the rest of the lattice have a smaller bond length[23]. Finally, an analysis of the data for p-type doping of GaAs:Ge (Ref. 24) and for Si:B (as described in Ref. 25) and Si:Ga (Ref. 26) support the negative magnitude for those two semiconductors. Accordingly, in our analysis, we favored the theoretical determination of a_v by of Cardona and Christensen[26] over those of Van De Walle[29]. In this work, a_c was obtained by adding this a_v to the experimentally determined a_{cv}, since excellent data based on diamond anvil cell experiments are now becoming available[30]. But, it should be noted that the recent determinations of a_c by Cargill, et al (Ref 27) and of a_v by Nolte, et al.(Ref. 28), support a positive sign for a_v in Si.

Band Gaps

We choose the $Al_xGa_{1-x}As$ system for our study. That system has the best characterized set of parameters. It is usual for the temperature dependence of each band gap to be expressed using the Varshni[31] form as,

$$\Delta E_{cv}^{\Gamma}(T) = \Delta E_{cv}^{\Gamma}(0) - \frac{\alpha^{\Gamma} T^2}{(\beta+T)},$$

where α^i and β are the so-called Varshni coefficients for the i^{th} band gap (β is taken as independent of band gap). While β is associated with the Debye temperature[6] and does not vary in a given semiconductor with the symmetry of the band gap, α does depend on the symmetry of the band gap involved (see discussion in Ref. 32). However, since we can write

$$-\frac{d\Delta E_{cv}}{dT} = \Delta S_{cv} = \alpha\left[1 - \left(\frac{\beta}{(\beta+T)}\right)^2\right],$$

and since it is a reasonable first assumption[7] that ΔS_{cv}(elec-phon) is independent of conduction band symmetry (since holes are much more effective at softening the lattice than an electron), we can subtract out the electron-phonon contribution from two different bands in the same semiconductor to obtain:

$$\begin{aligned}\Delta S_{cv}^{X} - \Delta S_{cv}^{\Gamma} &= 3\alpha_p\,(a_c^{\Gamma} - a_c^{X}) \\ &= (\alpha^{\Gamma} - \alpha^{X})\left[1 - \left(\frac{\beta}{(\beta+T)}\right)^2\right] \qquad (28)\end{aligned}$$

Above the Debye temperature, we see that

$$3\alpha_p^{(\text{high T})}\,(a_c^{\Gamma} - a_c^{X}) = (\alpha^{\Gamma} - \alpha^{X}) \qquad (29),$$

which is a remarkable relationship connecting the difference in the Varshni α coefficients to the difference in the conduction band deformation potentials. Figure 3 investigates these and other considerations for the $Al_xGa_{1-x}As$ system. Here we plot the total entropy for the Γ and X band gaps in $Al_xGa_{1-x}As$ as a function of Al composition, x. By subtracting off the thermal expansion component, we are left with the electron-phonon contribution alone, that is then fitted to a second order polynomial. Our assumption that $\Delta S_{cv}^{(\text{elec-phon})}$ is independent of band gap is supported for AlAs and for Al compositions above the Γ - X crossing, but some discrepancy arises for low Al compositions and GaAs. It appears that $\Delta S_{cv}^{(\text{elec-phon})}$ is larger for X band transitions than for Γ band transitions in GaAs. In the bond charge model, since both transitions involve removal of charge from the valence bond, this difference must be due to the different nature of the conduction band antibonding states. Furthermore, TA phonons are assumed to be the primarily responsible for the increase in entropy with increasing temperature[14]. Since TA phonons do not couple to spherically

symmetric electron states, this model would predict almost no $S_e^{(elec\text{-}phon)}$ for Γ_6 electrons. The experimental evidence at the GaAs end of the $Al_xGa_{1-x}As$ system supports this hypothesis; the Γ_6 electron states in GaAs are less effective than the $X_{6,7}$ states in softening the lattice. Thus we propose that while Eq. (10) holds for $X_{6,7}$ electrons (as one might expect since it was measured for GaP), the factor appears even larger for Γ_6 electrons. Again, however, we must caution that in AlAs, it appears that Eq. (10) holds for both conduction bands. Finally, there also is some evidence supporting a bowing of $\Delta S_{cv}^{(elec\text{-}phon)}$ in the alloy, but it might be equally well attributed to bowing due to the various parameters contributing to $\Delta S_{cv}^{(therm\ exp)}$. Table I details the sources for the parameters used.

Also not often noted is that because of the different deformation potentials associated with the different band gaps, $Al_xGa_{1-x}As$ will have a discontinuity in the temperature dependence of the bandgap at the direct-indirect crossover. This is in contrast to the linear interpolation typically applied to this system[33].

Band Offsets

The band offsets for the case of the pure binary compounds, AlAs/GaAs, were also studied. We assumed Eq. (10) to hold for all temperatures in all positive gap semiconductors, primarily because of the lack of other data. Here the effects of the thermal expansion component alter our previous conclusions in two ways. For any direct gap semiconductor, the large a_c means thermal expansion effects significantly contribute to the temperature dependence of the conduction band. As Eq. (24) above shows, a large a_c (or V_e) boosts the standard entropy of the electron. In fact, as Fig. 4 shows, the ratio S_e/S_h, or equivalently $(E_c(0K)-E_c(T))/(E_v(0K)-E_v(T))$ approaches unity in a direct gap semiconductor like GaAs. However, for AlAs, the small deformation potentials associated with the band edges places the ratios closer to the values described by Eq. (10) above, with the valence band dominating the effects as shown in Fig. 5.

Figures 6 and 7 show the temperature dependence of the band offsets between AlAs and GaAs. Now we see that while $\Delta E_v^{AlAs/GaAs}$ shows very little temperature variation, $\Delta E_c^{AlAs/GaAs}$ varies by 30meV from 0K to room temperature.

Conclusions

We have extended our previous considerations of the temperature dependence of band offsets by including thermal expansion effects. We found that whereas holes and the valence band produced the largest electron-phonon contribution, thermal expansion effects are primarily determined by the symmetry of the conduction band. For a Γ conduction band, thermal expansion effects make the overall electron entropy similar in size to holes in the valence band. For X minima, the hole entropy again dominates, while L minima fall somewhere between. While the magnitude of the effects predicted are relatively small for the semiconductors considered here and may be ignored in most cases, they are comparable to the overall band gap temperature dependences, and therefore, need

to be considered when designing any device operating at more than one fixed temperature.

Acknowledgements

This work was supported by the Air Force Office of Scientific Research. At the University of New Mexico, work was supported by contract number F49620-89-C-0028 while the work at Oregon State University was supported by grant number AFOSR-89-0309.

REFERENCES

[1]K.J. Malloy and J.A. Van Vechten, Appl. Phys. Lett. **54**, 937 (1989).

[2]J.A. Van Vechten and K.J. Malloy, J. Phys: Condens. Matter **1**, 91 (1990).

[3]W. Weber, Phys. Rev. Lett. **33**, 371 (1974), and Phys. Rev. B 15, 4789 (1977).

[4]R.M. Martin, Phys. Rev. **186**, 871 (1969), and Phys. Rev. B **1**, 4005 (1970).

[5]H. Brooks, Advan. in Electron. and Electron Phys. **7**, 85 (1957)

[6]C.D. Thurmond, J. Electrochem Soc. **122**, 1133 (1975).

[7]J.A. Van Vechten, in "Handbook on Semiconducors, Volume 3," edited by S. P. Keller (North Holland, 1980)

[8]C. Herring and M.H. Nichols, Rev. Mod. Phys. **21**, 185 (1949)

[9]M.J. Caldas, A. Fazzio and A. Zunger, Appl. Phys. Lett. **45**, 671 (1984).

[10]J. M. Langer and H. Heinrich, Phys. Rev. Lett. **55**, 1414 (1985).

[11]J. Tersoff and W. A Harrison, Phys. Rev. Lett. **58**, 2367 (1987).

[12]V. Heine and C.H. Henry, Phys. Rev. B **11**, 3795 (1975).

[13]P.B. Allen and M. Cardona, Phys. Rev. B **23**, 1495 (1981) and Phys. Rev. B **27**, 4760 (1983).

[14]V. Heine and J.A. Van Vechten, Phys. Rev. B **13**, 1633 (1976)

[15]W. Paul, J. Phys. Chem. Solids **8**, 196 (1959).

[16]J. Bardeen and W. Shockley, Phys Rev. **80**, 72 (1950).

[17]W.B. Gauster and D.H. Habing, Phys. Rev. Lett. **18**, 1058 (1967).

[18]W.B. Gauster, Phys. Rev. **187**, 1035 (1969).

[19]J.A. Vergés, D. Glötzel, M. Cardona, and O.K. Andersen, phys. stat. sol. (b) **113**, 519 (1982).

[20]M. Cardona and N.E. Christensen, Phys. Rev. B **35**, 6182 (1987).

[21]G.A. Samara and C.E. Barnes, Phys. Rev. B **35**, 7575 (1987).

[22]G.A. Samara, Phys. Rev. B **36**, 4841 (1987) and Phys. Rev. B **37**, 8523 (1988).

[23]U. Pietsch and K. Unger, phys. stat. sol. (a) **80**, 165 (1983).

[24]D.L. Rode, R.L. Brown, and M.A. Afromowitz, Journal of Crystal Growth 30, 299 (1975).

[25]"Landolt-Börnstein Numerical Data and Functional relationships in Science and Technology," edited by O. Madelung, M. Schultz and H. Weiss (Springer, Berlin, 1982) Vols. 17a, 17b and 22.

[26]M. Cardona and N.E. Christensen, Phys. Rev. B **35**, 6182 (1987).

[27]C.S. Cargill, III, J. Angilello, and K.L. Kavanagh, Phys. Rev. Lett., **61**, 1748 (1988).

[28]D.D. Nolte, W. Walukiewicz, and E.E. Haller, Phys. Rev. B **36**, 9392 (1987).

[29]C.G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).

[30]D. J. Wolford and J. A. Bradley, Bull. Am. Phys. Soc. **29**, 291 (1984).

[31]Y. P. Varshni, Physica **34**, 149 (1967).

[32]J. S. Blakemore, J. Appl. Phys. **53**, R123 (1982).

[33]S. Adachi, J. Appl. Phys. **58**, R1 (1985).

[34]B. Monemar, Phys. Rev. B **8**, 5711 (1973).

[35]H.C. Casey and M. Panish, in "Heterostructure Lasers, Part B: Materials and Operating Characteristics," (Academic Press, New York, 1978), p. 9.

[36]H. Shen, S.H. Pan, Z. Hang, J. Leng, F.H. Pollack, J. M. Woodall, and R.N. Sacks, Appl. Phys. Lett. **53**, 1080 (1988).

[37]H. Neumann and W. Junge, phys. stat. sol. (a) **34**, K39 (1976).

Symbol	Variable Name	GaAs Value	AlAs Value
α^{Γ}	Varshni Coeff	5.41×10^{-4} eVK^{-1} Ref. 32	7.87×10^{-4} eVK^{-1} fit to Ref. 34
α^{X}	Varshni Coeff	4.60×10^{-4} eVK^{-1} Ref. 32	6.0×10^{-4} eVK^{-1} Ref. 35
β	Varshni Coeff	204 K Ref. 32	408K Ref. 35
$\alpha_P(T)$	Thermal Expan coefficient	Ref. 32	Ref. 25
β_T	Isothermal Bulk Modulus	78.9 GPa Ref. 32	78.1 GPa Ref. 25
$\Delta V_{cv}(\Gamma)$	Standard Volume	10.74 meVkbar^{-1} Ref. 30	10.2 meVkbar^{-1} Ref. 33
$\Delta V_{cv}(X)$	Standard Volume	-1.34meVkbar^{-1} Ref. 30	-1.34meVkbar^{-1} Set equal to GaAs
a_v	Deformation Potential	-1.6 eV Ref. 20	-1.2 eV Ref. 20
a_c^{Γ}	Deformation Potential	$a_c^{\Gamma}=a_{cv}^{\Gamma} + a_v$	$a_c^{\Gamma}=a_{cv}^{\Gamma}+a_v$
a_c^{X}	Deformation Potential	$a_c^{x}=a_{cv}^{x} + a_v$	$a_c^{x}=a_{cv}^{x} + a_v$

TABLE I:

This Table indicates the value and sources of the parameters for the calculations used in Figures 3-6.

FIGURE CAPTIONS:

Figure 1. A qualitative plot of the standard free energies of formation referred to in Equations 1-4 for a conventional semiconductor such as GaAs or AlAs. The vacuum level serves as the reference level here and we note that Eq. (2a) has a positive standard free energy while Eq. (3) has a negative standard free energy.

Figure 2. A plot of the standard volume of formation for a Γ electron- hole pair as a function of the bond length of the III-V compound semiconductor. The dotted line is a fit to the data, which were obtained from Ref. 25.

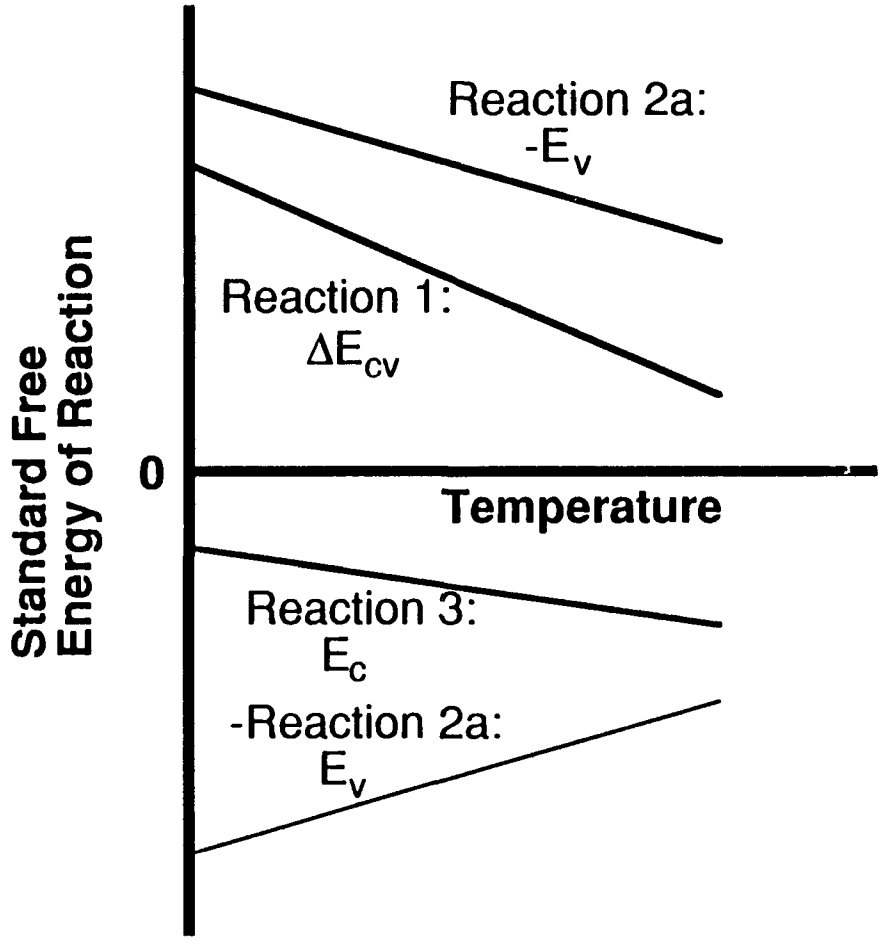
Figure 3. The Standard Entropy of Formation for both Γ_6 and X electrons- with Γ_8 holes in the $Al_xGa_{1-x}As$ system. Thermal expansion components are then subtracted off to obtain a measure of $\Delta S_{cv}^{(electron\text{-}phonon)}$, which are then fitted by second-order polynomials. Data points were obtained from Ref. 32 for GaAs X and Γ, Ref. 36 for $Al_{0.18}Ga_{0.82}As$, Ref. 37 for $Al_xGa_{1-x}As$ for $x > 0.4$, and from Refs. 34 and 35 for AlAs.

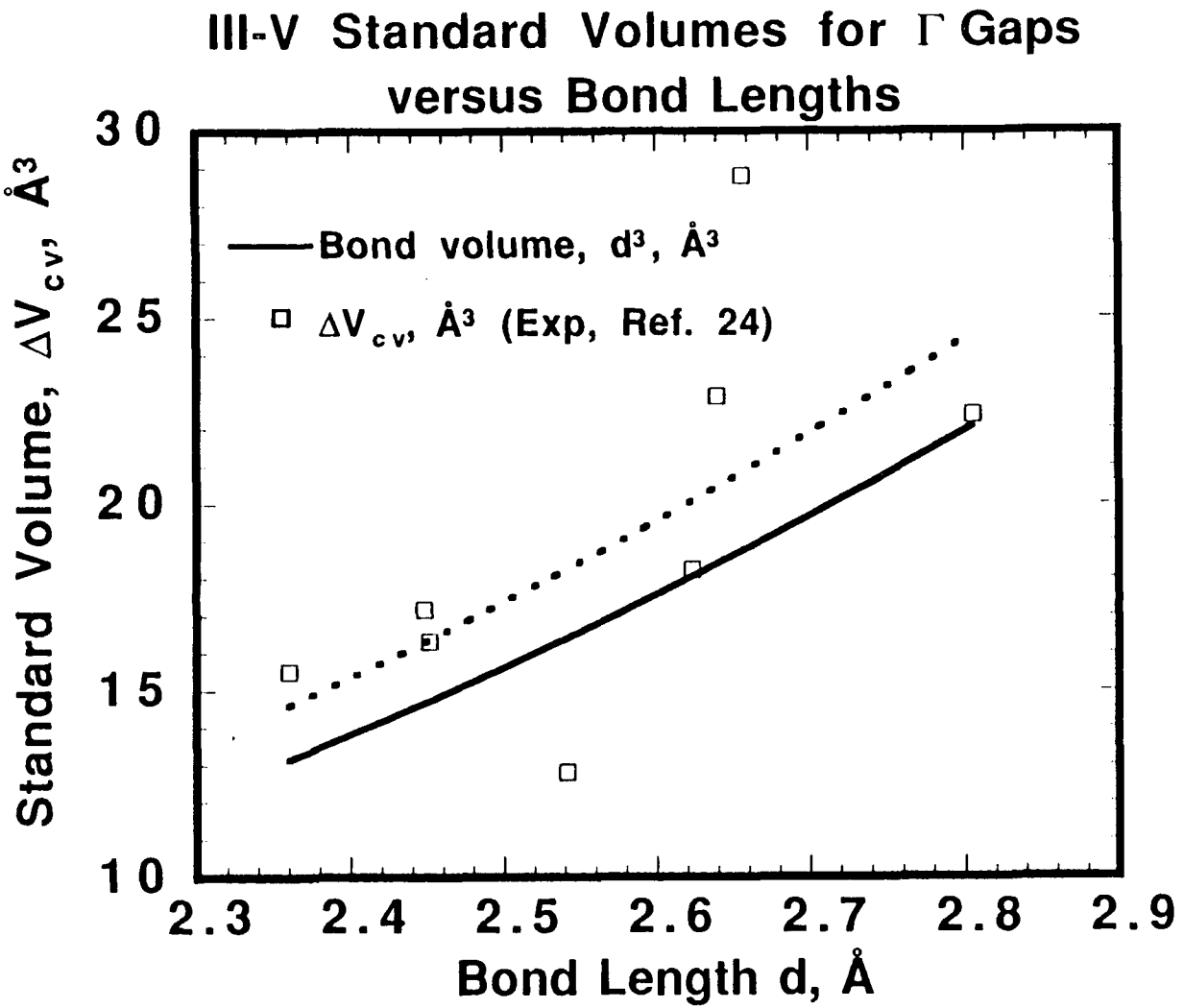
Figure 4. Using the procedure outlined in the text, the temperature variation of the individual band edges of GaAs were obtained. Note that the conduction band and valence band's change with temperature are nearly equivalent.

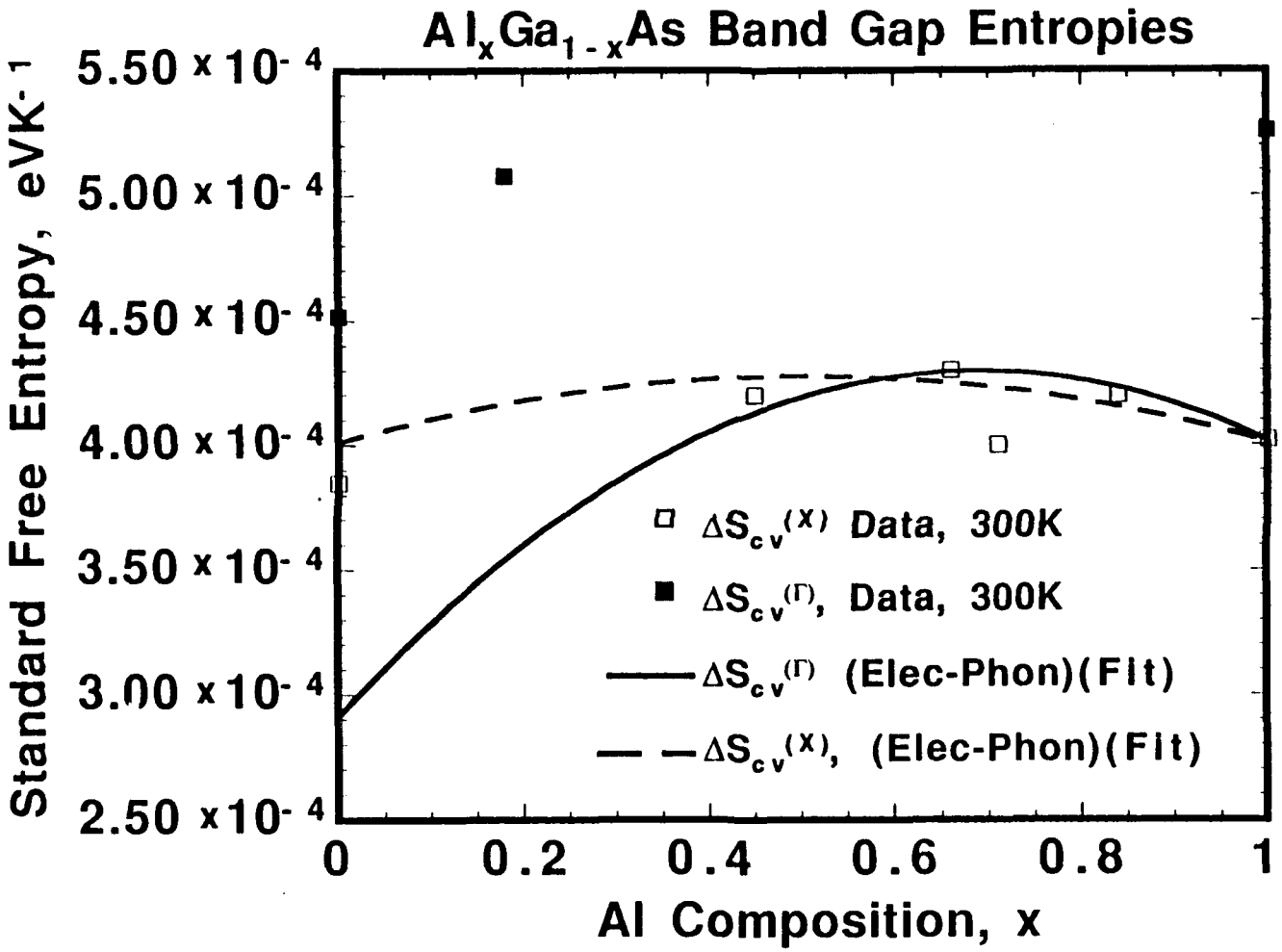
Figure 5. Same as Fig. 3 except for AlAs. Now the valence band shift is much greater than the conduction band shift.

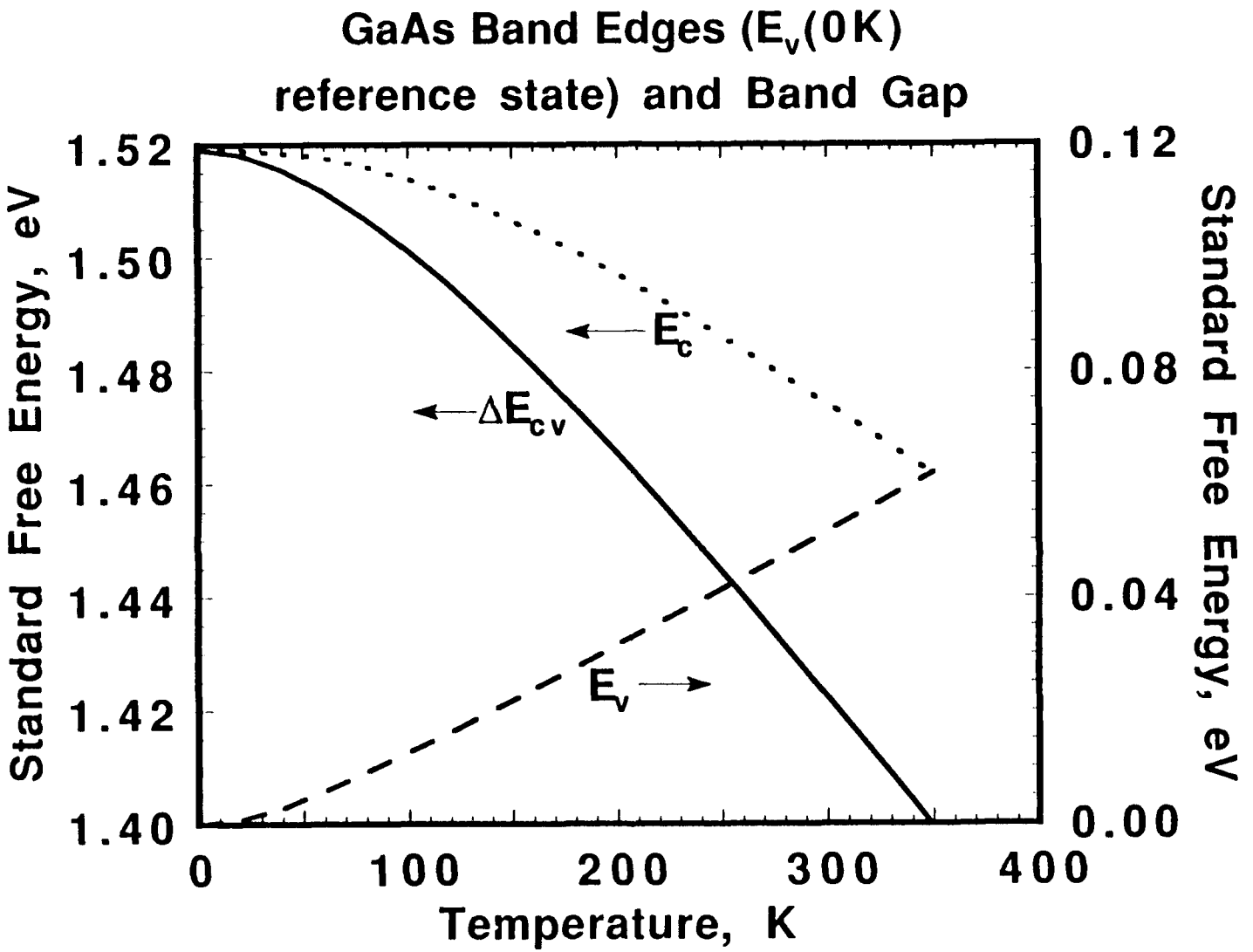
Figure 6. The valence band offset variation for GaAs/AlAs with temperature.

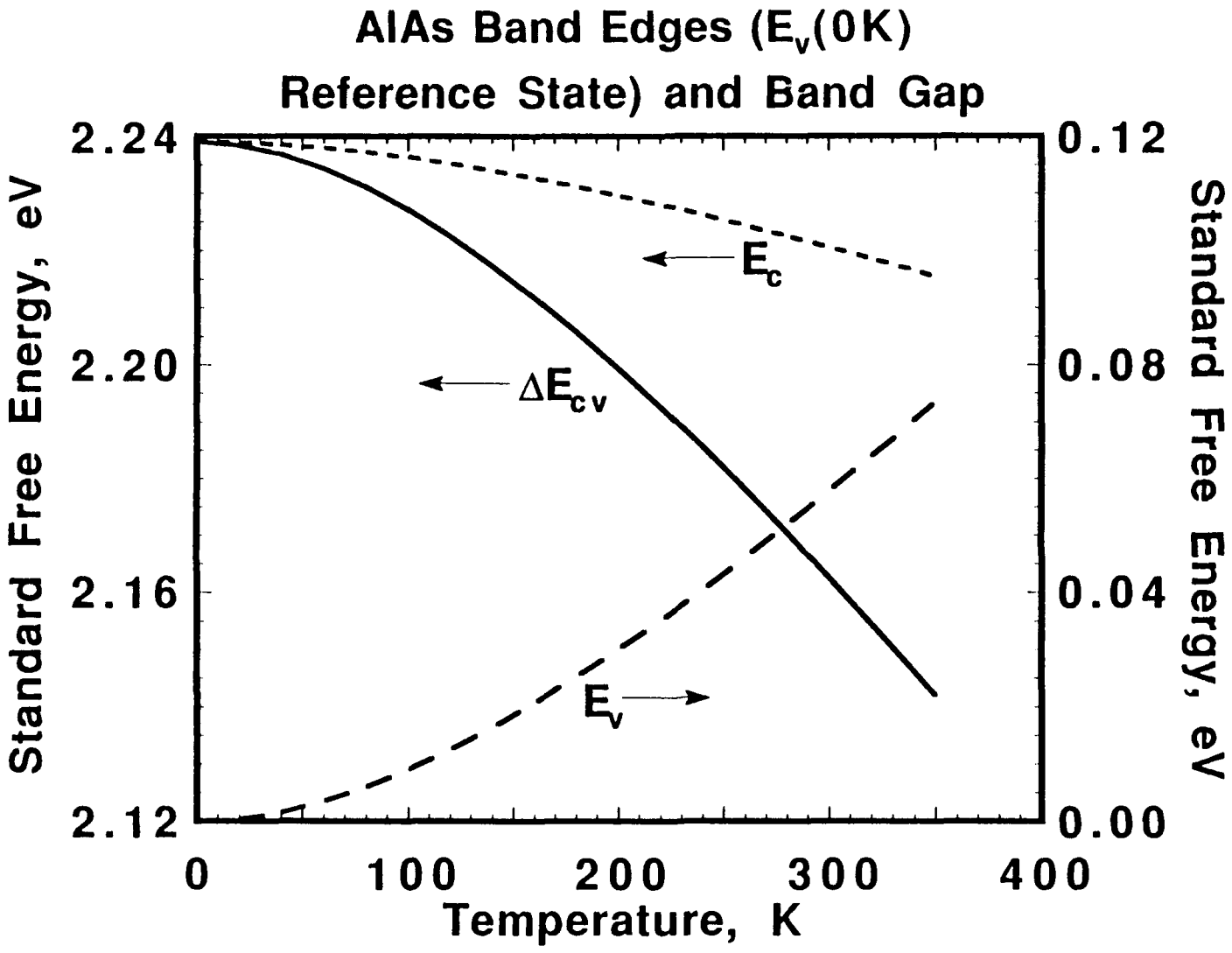
Figure 7. The conduction band offset variation for GaAs/AlAs with temperature.

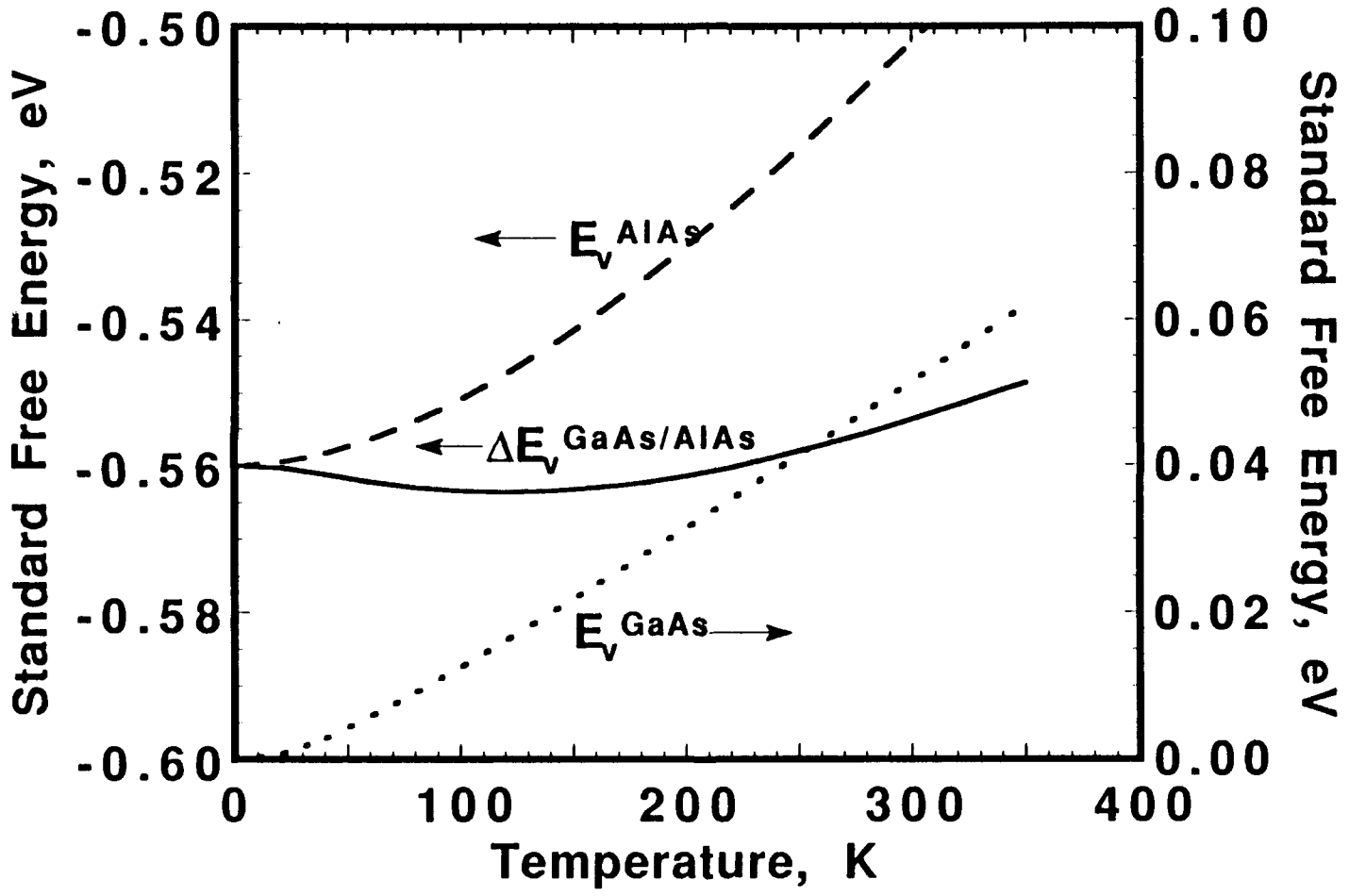

Reaction 2a:
$-E_v$
Reaction 1:
ΔE_{cv}
0
Temperature
Standard Free
Energy of Reaction
Reaction 3:
E_c
-Reaction 2a:
E_v

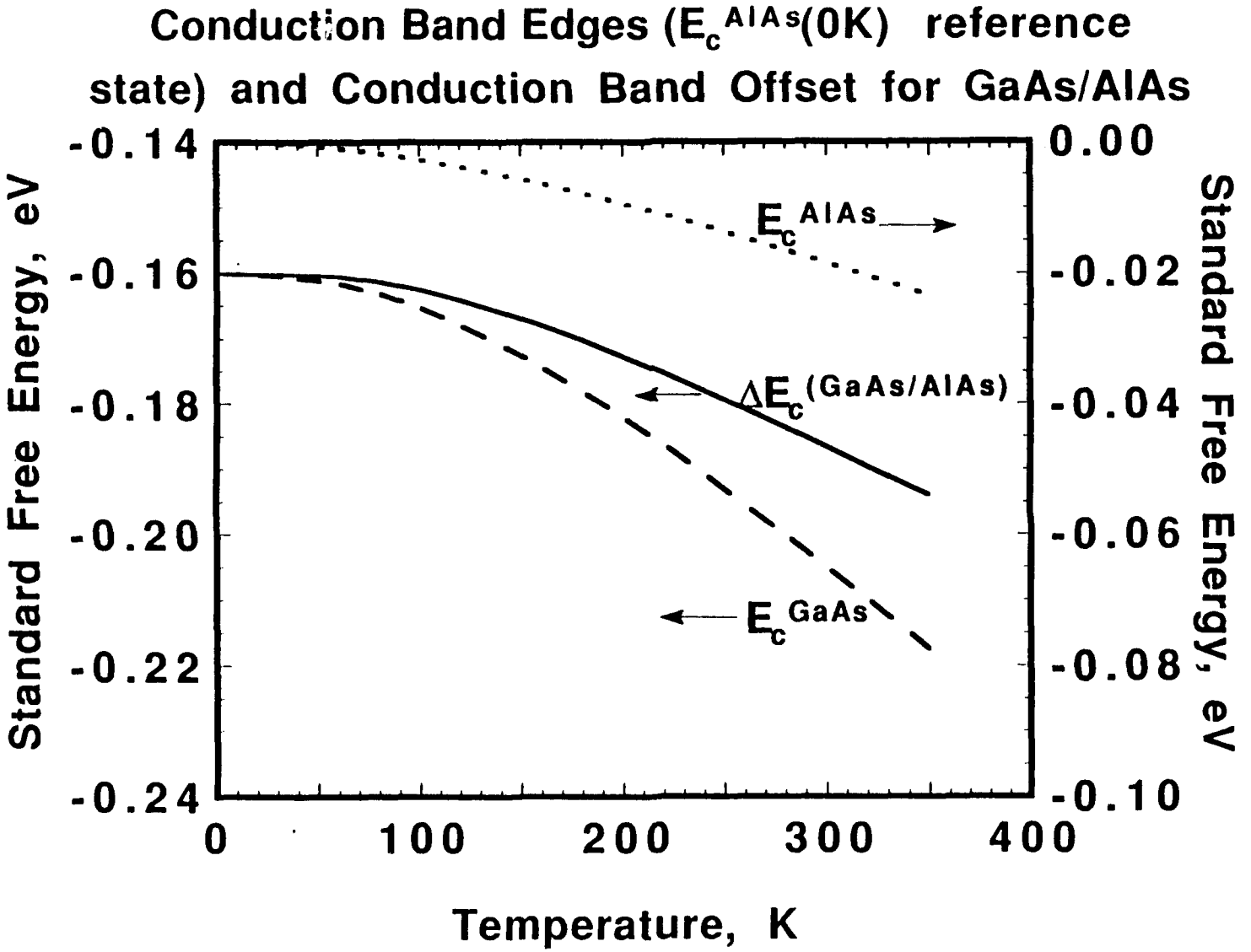

III-V Standard Volumes for Γ Gaps
versus Bond Lengths
Standard Volume, ΔV_cv, Å3
30
25
20
15
10
Bond volume, d3, Å3
ΔV_cv, Å3 (Exp, Ref. 24)
2.3
2.4
2.5
2.6
2.7
2.8
2.9
Bond Length d, Å


Al$_x$Ga$_{1-x}$As Band Gap Entropies
Standard Free Entropy, eVK^{-1}
5.50 x 10^{-4}
5.00 x 10^{-4}
4.50 x 10^{-4}
4.00 x 10^{-4}
3.50 x 10^{-4}
3.00 x 10^{-4}
2.50 x 10^{-4}
0
0.2
0.4
0.6
0.8
1
Al Composition, x
$\Delta S_{cv}^{(X)}$ Data, 300K
$\Delta S_{cv}^{(\Gamma)}$, Data, 300K
$\Delta S_{cv}^{(\Gamma)}$ (Elec-Phon)(Fit)
$\Delta S_{cv}^{(X)}$, (Elec-Phon)(Fit)


GaAs Band Edges (E_v(0K) reference state) and Band Gap
Standard Free Energy, eV
1.52
1.50
1.48
1.46
1.44
1.42
1.40
0.12
0.08
0.04
0.00
Standard Free Energy, eV
E_c
ΔE_{cv}
E_v
0
100
200
300
400
Temperature, K


AlAs Band Edges (E_v(0K) Reference State) and Band Gap
Standard Free Energy, eV
2.24
2.20
2.16
2.12
0.12
0.08
0.04
0.00
Standard Free Energy, eV
E_c
ΔE_{cv}
E_v
0
100
200
300
400
Temperature, K


Valence Band Edges (E_v^{GaAs}(0K) reference state) and Valence Band Offset for GaAs/AlAs
Standard Free Energy, eV
-0.50
-0.52
-0.54
-0.56
-0.58
-0.60
0.10
0.08
0.06
0.04
0.02
0.00
Standard Free Energy, eV
E_v^{AlAs}
$\Delta E_v^{GaAs/AlAs}$
E_v^{GaAs}
0
100
200
300
400
Temperature, K


Conduction Band Edges (E_c^{AlAs}(0K) reference state) and Conduction Band Offset for GaAs/AlAs
Standard Free Energy, eV
-0.14
-0.16
-0.18
-0.20
-0.22
-0.24
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
Standard Free Energy, eV
E_c^{AlAs}
ΔE_c(GaAs/AlAs)
E_c^{GaAs}
0
100
200
300
400
Temperature, K

Defect Control in Semiconductors
K. Sumino (ed.)
© Elsevier Science Publishers B.V. (North-Holland), 1990

RESOLVING DEFECT MEGA-CONTROVERSY BY GIGA-EVENT MONTE CARLO SIMULATION OF THE MACRO-CONSEQUENCES OF ATOM-LEVEL ASSUMPTIONS ON MICROCOMPUTERS AT NANO-COST*

J. A. Van Vechten, U. Schmid† and N.C. Myers‡

Center for Advanced Materials Research, Electrical and Computer Engineering Department, Oregon State University, Corvallis, OR 97331-3211 USA

After discussing reasons that controversy is so common in this field, we propose a mechanism for resolving it at very low cost and in reasonable time. We present VIDSIM, a program that simulates complex defect processes in background on PCs with user specified assumptions. Rates as great as those for supercomputers are obtained by running several PCs. We illustrate its utility by showing that the consequences of the "kick-out" mechanism for Au diffusion into Si are not at all those attributed to it in recent literature.

1. INTRODUCTION

1.1. Origins of controversy

As this Conference has made perfectly clear, control of defects is critical to the progress of modern electronics. Thus, many people of diverse background, intuition and mode of work try to contribute to this science and art.

Most of us begin our attempts by studying mathematics, physics, and other material sciences. We are convinced of the fundamental truth of, e.g., the Schrödinger equation, but we also learn that it has been solved exactly only for cases no more complex than the He atom; we are told that the finite mass of the universe and the Heisenberg Uncertainty Principle imply that memory sufficient to solve the U atom wave function exactly can not exist. Thus, we must make approximations and try out assumptions. However, mathematics teaches us two discouraging theorems. First, the Inconsistency Theorem teaches us that, if any inconsistency is allowed into any logical system, then any conclusion can be derived from any assumption. (When challenged by a student to prove that, given 1 = 2, he was the Pope, the great British mathematician Bertrand Russell replied "The Pope and I are two, tis true, but if two be one, then the Pope and I are one.") Thus, the uncontrolled approximations that we are forced to make introduce an ever present and unpredictable potential for massive error. Second, the Incompleteness Theorem teaches us that our science must contain many true statements which can never be proved in the rigorous, mathematical sense. Evidently, many valid assumptions will never be verified.

Meanwhile, the semiconductor industry proceeds with a great many "cut and try" experiments and a vast literature of empirical observations is generated. In all practical circumstances, the problem of the control of defects in semiconductor devices is a very complex subject and in many cases there are contradictory reports. Major controversies rage re almost every significant semiconductor processes and product. These include the roles of vacancies and of self-interstitials in atomic diffusion in Si and the nature of the self-compensating EL2 and DX centers in GaAs or AlGaAs. The consequences for reliability and yield of large misfit strains being designed into Si epi-CMOS, Si-Ge, and III-V heterostructure devices are

*Supported in part by U.S. Air Force Office of Scientific Research under contract AFOSR-89-0309.

†Present Address: Max Planck Institut Für Festkörperforschung, Stuttgart, FRG.

‡Present Address: Mentor Graphics, Inc., Beaverton, OR.

not agreed upon. The subject is characterized by various assertions of simplifying assumptions and appeals to intuition. However, different people have different intuitions as well as differing opinions re the appropriate approximation. It is said that only two empirical laws are well established: a) All experiments measure something; and b) All theories are wrong, it is just a matter of degree.

1.2. Ab initio methods on supercomputers

Some people try to make progress with ab initio methods on supercomputers. These are very expensive machines and highly specialized skills are required to use them. They are often used to make "fundamental" studies of defects in what is regarded to be a "state of the art" approximation, such as the local density functional approximation, LDA.[1-4]

Despite the availability of about 100 supercomputers world-wide (e.g., the Cray XMP or the IBM 3090 VAP), these efforts have not sufficed to resolve difficulties that cost large sums of money. Let us consider why this might be so for the particular case of the LDA.

The LDA has demonstrated a good ability to predict bulk ground state properties, such as the lattice constant and the bulk modulus.[5,6] However, the LDA is not accurate for quantities, like the band gap, which involve excited electronic states; e.g., it predicts the band gap of Ge and GaAs to be zero.[7]

It has also been established that when electron correlation and exchange are treated in an even better approximation (time dependent screened Hartree-Fock) than the LDA, a reasonable band gap is obtained.[8,9] However, these better calculations are beyond the state of the art even with the present generation of supercomputers for the treatment of interesting defects in imperfect hosts. Also, Hanke and coworkers have shown that the problem with the LDA that leads it to calculate band gaps incorrectly is equivalent to an incorrect treatment of the local electrostatic fields, LEF, within each unit cell of the host crystal.[10-12] By LEF we mean the electrostatic field that would act on an ideal, point test charge. The LEF varies within the unit cell with spatial components of all reciprocal lattice vectors and varies in time with the spectrum of all virtual transitions allowed in the sample, as well as with the space and time variation of any applied field.[13-15]

It is common practice to correct the band gaps calculated with the LDA by simply displacing the band extrema as needed to agree with experiment (i.e., to use the "scissors operator"). However, one should worry that the correction required for the effect of the true LEF on the band edge states, which are delocalized throughout the sample, might be entirely different from those LEF corrections appropriate to the tightly localized electronic states associated with the defects of interest here. This point has been made before.[11] The worry should be particularly great when one uses LDA methods to examine problems relevant to this Conference, e.g., the hypotheses that the EL2 compensation mechanism in GaAs occurs because of a displacement of an ionized and isolated As on Ga antisite defect, As^{+}_{Ga}, from the ideal lattice Ga site[16] or that the DX self compensation mechanism in AlGaAs results from a displacement of isolated donors.[17]

Positron annihilation studies, such as those presented[18,19] by Dannefaer et al., serve to illustrate the problem of the LDA mistreatment of the LEF. The most complete, first principle, LDA calculations of positron properties in Si and GaAs are probably those of Puska et al.[20] They calculate annihilation lifetimes for positrons trapped in single vacancies in reasonable agreement with experiment; for Si they calculate 254 to 261 ps depending on charge state while the empirical value[18] is about 270 ps and for neutral Ga and As vacancies in GaAs they calculate 267 and 279 ps compared with an empirical value of about 260 ps that is about the same for either.[19] This may be taken to indicate that the LDA calculation obtained the

charge density in the unit cell and about the vacancies rather accurately. However, as with the random phase approximation in general, the LDA method does not get the LEF properly from this charge distribution.[13-15] Because positron experiments are done with only one positron in the sample at a time, and because this positron need not be orthogonalized to any of the electron wave functions of the sample, the positron is an ideal probe of the LEF around the vacancy cavity where it is trapped. The energy required to remove the positron from the center of the vacancy cavity to the interstitial channel of the perfect host crystal is just the integral of the LEF over this path and is called the positron binding energy, E_b. Calculated values for E_b in Si range from 1.6 eV for V^{2-} to 1.0 eV for V^0 to 0.5 eV for V^{2+} while experiment[15] indicates that all values exceed 3 eV. A similar discrepancy occurs for single vacancies in GaAs.

We conclude that the LDA methods are quite incapable of determining either the ionization levels for deep level defects or the defect energy of formation as a function of atomic configuration for charged states. These energies and the activation energies of metastable configurations related to them have a scale of ~ 2 eV while the improper treatment of the LEF introduces errors of order 2 eV.

Therefore, it seems that practical progress will be made only with semi-empirical theories and with experiment for the foreseeable future.

1.3 Atom-level versus macroscopic defect information

Information re defects in semiconductors tends to divide into two classes: a) atom-level information from optical spectroscopy, spin spectroscopy,[21,22] capacitance transient spectroscopy,[23] positron annihilation,[18,19] Mössbauer,[24,25] etc; and b) macro information re diffusion and other defect profiles, free carrier distributions, lifetime profiles, etc., i.e., the parameters directly relevant to practical devices. We are here concerned with the relation between these two classes. Our position is that the latter should be understood in terms of the former through the application of thermodynamic reasoning.

The task of making the proper connection between the atom level properties of known defects and the macro consequences relevant to practical devices is very complex. If we take the case of Si and the established properties of vacancies therein, V_{Si}, we note that V_{Si} is known[21,22] to: a) have five ionization states within the band gap; b) exhibit negative-U properties with respect to its positive (donor) ionization states; c) exhibit large, symmetry breaking, Jahn-Teller distortions which imply a complex elastic field; d) exhibit recombination enhanced diffusion at low temperatures, T; e) form divacancies with a large binding energy that has a long range component as well as a nearest neighbor component.[26] Although we, and a few others, have tried to deduce the net diffusivity via vacancy hopping processes for Si and for impurities in Si from the established atom-level properties of vacancies in Si,[26-31] all such work contains approximations, appeals to intuition, imposed boundary conditions, passages to limits, hypotheses, applications of a selection of results that hold for equilibrium to processes, such as diffusion, that are not equilibrium, etc., etc. The controversy that exists in this literature illustrates the Inconsistency Theorem. We have concluded that only "experimental theory", i.e., Monte Carlo simulations, can clarify these issues.

When considering atomic diffusion processes via V_{Si}, several workers have sought to avoid the complication of the thermal distribution V_{Si} in their various ionization states with differing activation energies for strictly thermal hopping and for hopping enhanced by the recombination of the thermally generated electron-hole pairs,[29] with their coulomb and elastic interactions with atoms and other defects, etc., etc., by assuming the totality of vacancy mediated diffusion can be described by one,

Fermi level dependent diffusivity function.[31-33] This function is parameterized with some selected macro data. The method amounts to trying to force a Fickian diffusion formulation upon the problem despite the fact that it is well known that diffusion profiles in Si and other semiconductors deviate markedly from the simple forms predicted by Fick's Laws. Recall that Fick's formulation of diffusion rests on three assumptions: a) the things that diffuse do not change during the process; b) the hops are uncorrelated one to the next; and c) the hopping probability is local, i.e., determined by circumstances in the first few shells of nearest neighbors about the thing that will hop with no long range effects such as are produced by long range coulomb and elastic fields. While these assumptions may be reasonable for the case of metals and plausible for the case of insulators, none of the three is applicable for the semiconductors of interest. Hence, it should be no surprise that observed diffusion profiles are not as predicted with a simple one component Fickian model.

Upon noting that their particular one component Fickian model for vacancy mediated atomic diffusion fails to account for empirical data, some workers propose that the reasonable conclusion is that another point defect, a self-interstitial, Si_i, must also be present in concentrations sufficient to mediate a major part of the total diffusivity that is observed.[31-33] However, the information re Si_i and other host and impurity interstitials obtained from atom-level experiments implies that they are much too unstable, i.e., have too high energies of formation, to be important contributors to thermal diffusion.[21,22,27,34,35] For this reason, in the signal paper re this approach Seeger and Chik proposed[36] that the self-interstitial contributing to thermal diffusion at T >> 300 K is nothing like that produced by irradiation in many atom-level experiments,[21,22] but instead is a sort of amorphous "extended interstitial" that is supposed to be stable only at high T and only on account of a high entropy. While this seems to be a still viable hypothesis,[37] most experimental and all theoretical treatments in the recent literature discuss the self-interstitial as if it were the simple interstitial, Si_i, as produced by irradiation at a cost of 12 eV displacement energy.

In our view this invocation of thermal distributions of Si_i with parameters adjusted to empirical macro data in some way is nothing more than a demon thesis and ought not to be allowed until after a proper treatment of the consequences of the various vacancies and vacancy complexes which are known to be present at processing temperatures.[18] Observations of the growth of extrinsic dislocation loops is often claimed to be convincing evidence for Si_i, but that claim was discredited[38] long ago. Effects of oxidation upon the rate of diffusion of dopants has also recently been claimed to demonstrate that oxidation injects Si_i but studies of the oxidation of Si-Ge alloys show the Ge to be snow-plowed in front of the oxidation front and not distributed either into the oxide or into the substrate; this demonstrates that the dominant direction of Si diffusion is from the bulk to the oxidation front, which implies that oxidation injects vacancies or vacancy complexes - not interstitials.[39]

However, we must admit there are several groups following Si_i approach. We must also admit that the proper treatment of vacancy mediated diffusion, or of interstitial mediated diffusion, will require a huge amount of computing that until now has not been feasible. It is to make this task feasible that we have developed, and continue to enhance, a computer program, VIDSIM = vacancy and interstitial diffusion simulator,[40-42] which allows virtually anyone interested in the subject to test the macro consequences of whatever atom-level assumptions he wishes, in background mode, on one or more microcomputer at almost no cost. We will provide VIDSIM to anyone requesting it at the cost of the media plus postage.

1.4 VIDSIM experimental theory

Our approach with VIDSIM to the problem of establishing the relation between atom-level assumptions and macro consequences is simple brute force - a computer experiment on each theory. One inputs a set of assumptions adequate for the microcomputer, PC, or other computer, to determine the activation free energy of any process that one wants to allow, an initial configuration, T, and a seed for the random number generator. The PC tabulates the mean time for anything that can happen at any point in the present configuration where something can happen (site of a mobile defect). The sum of the inverse of these times is the inverse of the net mean time for something to happen somewhere. Using the Mean Value Theorem, VIDSIM assumes each step occurs after the net mean time calculated in this way. Possible events are assigned a section of the 0 to 1 probability line segment according to their inverse mean times. The random number generator selects a number in the 0 to 1 interval and the particular event to occur after the net mean time is selected thereby. Then, all possible events affected by this event are reconsidered and the event table is altered accordingly, the new net mean time is calculated and the process is repeated as many times as may be necessary to generate reliable statistics re the macroscopic process of interest.

It seems that this brute force strategy has never been seriously attempted before because it requires so much computing. We demonstrate herein that it is now quite practical and costs almost nothing because of the dissemination of large numbers of PC's with very impressive computational power which is grossly underutilized worldwide. As of this writing we have simulated several billion events on several IBM PC/AT, IBM S2/70 and INTEL 302/25 PC's. With such a machine VIDSIM can simulate a sample as large as 1.4×10^{14} lattice sites with as many as 8600 defects of which 400 may be mobile (vacancies and interstitials). Depending on the simulation, VIDSIM will require 120 kb or more of the 640 kb of main memory available on such a machine running under DOS for a background simulation using commercial multitasking software such as DoubleDos or DESQview.This leaves adequate memory for most foreground programs. We find that, e.g., the IBM S2/70 with 20 MHz math coprocessor can run VIDSIM about 4% as fast as it runs on a CRAY XMP. With the mix of PCs we find idling around in our small university department, we have about 75% of a CRAY XMP available *all the time*. This means we have about 6570 XMP equivalent cpu hours per year available to run VIDSIM and we obtain this resource at no cost because the PCs were obtained on other justifications. We simply fill up the idle capacity of these machines. It would cost us about $\$6.5\times10^{6}$/year to do the equivalent calculation on an XMP. Of course, with an XMP, or any machine with a longer word length and more available RAM, we could simulate larger samples, but we do not believe this to be necessary.

1.5 Controversy re analytic treatment of atom assumptions

The alternative to this massive, but now practical, computing which almost everyone has previously taken is to deduce analytic expressions for the defect formation, migration and reaction from the atom-level properties one decides to ascribe to the defects one assumes to be present. As noted above this is a very complex problem for real defects in real semiconductors. As all such treatments have invoked approximations, assumptions and boundary conditions, different authors obtain different analytic expressions for ostensibly the same situation and there is much controversy.

A clear example of a controversy between authors attempting to deduce analytic expressions from the same atom-level assumptions, which we have chosen to examine with VIDSIM, is the diffusion of Au into Si according to the "Frank-Turnbull", FT, mechanism.[33,43-47] The FT mechanism is a highly simplified version of the

assumption that Au diffuses into Si only as an interstitial atom, Au_i, that it encounters V_{Si} and annihilates them to form substitutional Au, Au_{Si}. No account is taken of the charge states of V_{Si}, Au_{Si}, or of Au_i (which are unknown), of the recombination enhancement of V_{Si} diffusion (which may be dramatically affected by the Au_{Si} because this is such an effective lifetime killer), of the effect of variations of Fermi level due to surface band bending or Au profile, or of any elastic interactions. With the FT one assumes there are no Si_i and that there is no diffusion of Au via interchange between V_{Si} and Au_{Si} nearest neighbors.

Willoughby and co-workers concluded[44-46] that the FT was consistent with empirical observations that: a) the Au_{Si} profile from a two-sided diffusion is more "U shaped" than the efrc(x) predicted by the simple Fickian formulation; b) that an asymmetric U shaped profile is obtained from a one sided diffusion; and c) that the concentration of Au_{Si}, $[Au_{Si}](t)$, increases with time, t, roughly as $t^{1/2}$. However, Seeger and co-workers concluded[33,47] that the FT cannot conform with observation a) or, in particular, to c); they conclude that FT must produce an $[Au_{Si}](t)$ that is linear in t. The ensuing controversy has raged for many years.

Having concluded that the FT could not account for observations re the diffusion of Au into Si, Seeger proposed[47] the "kick-out model", KO, to account for the same observations and deduced analytic expressions for diffusion of various elements via this mechanism. With the pure KO one assumes there are no V_{Si}. Au is again assumed to diffuse into Si as Au_i, but is assumed to displace host Si atoms to form Au_{Si}, the only form of Au defect actually observed. The Si_i then diffuses to a free surface or other sink for interstitials. It is argued that it is the out-diffusion of Si_i that limits the increase of $[Au_{Si}]$ and that it is the distinction between this and the in-diffusion of V_{Si} in the FT that accounts for the asserted difference in the functional form of $[Au_{Si}](t)$. Again the physics of the problem is greatly simplified by neglecting any charge state effects of either Si_i or Au_i, any Fermi level or recombination enhancement effects, and any elastic effects.

We have thoroughly examined the pure KO mechanism with VIDSIM and present our results below. We are forced to conclude that the macroscopic consequences of these atom-level assumptions are entirely different from those ascribed to them by Seeger and in more recent literature particularly re the three empirical features of the diffusion of Au into Si noted above.

Several authors have considered the combined effect of both vacancies and interstitials,[28-33] usually a sum of the FT and KO mechanisms particularly for the diffusion of dopant elements into Si. Here also controversy rages. Starting from essentially the same data and atom-level model three groups come to entirely different conclusions. Matiot and Pfister conclude that vacancy and interstitial mechanisms make about equal contributions[28] to dopant diffusion in Si, while Yoshida concludes[30] that interstitials make very little contribution and Fahey et al. conclude[31] interstitials make dominant contributions.

2. EXPERIMENTS ON THE KO THEORY

Space does not permit a thorough description of VIDSIM here. The reader is directed for this to Ref. 40 and 42 and to the program disk itself, which we are happy to supply at the cost of the 1.2 Mb disk plus postage. We provide source codes as well as the compiled programs, documentation, and instructions.

The disk contains the general program, VIDSIM.EXE, and a few special cases including two versions of the KO theory, SIM_KO.EXE and SIM_KI.EXE. These differ on the assumption how new Au_i are injected into the Si. SIM_KO.EXE assumes a new Au_i is injected when a Si_i reaches the surface, i.e., that the Au injection is "stimulated' by the arrival of Si_i; this keeps the total number of interstitials in the sample

constant. SIM_KI.EXE assumes that instead the injection of Au_i is "spontaneous", i.e., occurring randomly (at a constant average rate) on the surfaces; this allows the number of interstitials present to vary and seems to us to be the more reasonable assumption. We have found no discussion of this point in the KO literature and have investigated both extreme assumptions. As we find no major qualitative difference between the two, we suppose the true situation must lie between the two extremes and conclude that, in fact, it does not make much difference. Unless otherwise noted, the results presented here are for the "spontaneous", SIM_KI simulation. We assumed both Si_i and Au_i are bond-centered interstitials; we find not empirical evidence nor any discussion in the KO literature on this point and we are convinced that it makes no qualitative difference to our results. We account for the sample time, t, simulated in units of the attempt period of the lattice at T. Although not firmly established empirically, the attempt period is generally, and we think safely, assumed[27] to be of the order of the Debye period times an entropy factor, i.e., of order 1×10^{-13} sec. Uncertainty in this time unit has no qualitative effect on our conclusions.

Of course, a critical issue is the sensitivity of the KO result on the values of the parameters assumed in the simulation. We have investigated this in what we think is an exhaustive manner, which we now describe. (Any user can change these parameters however he/she wishes.) The temperature was fixed at T = 1368.5 K = 1095 C, a value used in typical experiments. The single bond enthalpy between two Si atoms was taken from the enthalpy of melting[27] to be 180 meV; that between two Au atoms was taken from the enthalpy of melting of Au to be 29 meV. (We have checked that the pure Si crystal melting point comes out correctly by simulating a sample with a free surface above and below the true melting point, T = 1685 K.) We set the bond enthalpy between Si and Au to 0 meV.

For his thesis,[40] Schmid concluded from the KO literature that the enthalpy of formation of Si_i should be 4.4 eV and its enthalpy of migration should be 0.4 eV. Following the Datareview[48] of Weber, we have taken the enthalpy of formation of Au_i to be 2.5 eV. We took its enthalpy of migration to be 2.2 eV rather than Weber's 0.39 eV; this is a concession to economy of simulation devised and examined in Ref 40. When one starts doing a KO simulation, one notices that the Si_i create a great many uninteresting exchange events,

$$Si_i + Si_{Si} \Longrightarrow Si_{Si} \quad Si_i \ , \qquad (1)$$

especially when random walking around in relatively Au-free bulk regions. Also the kick-out exchange events,

$$Au_i + Si_{Si} \Longleftrightarrow Au_{Si} + Si_i \ , \qquad (2)$$

are more interesting than the simple migration of either Si_i or Au_i from one interstitial site to another. To enhance the ratio of interesting to uninteresting events, we chose to: a) increase the migration enthalpy assumed for Au_i so as to increase the relative probability of (2) instead of a simple migration; and b) make the simulated samples much thinner than VIDSIM allows. With this change the probability that a Au_i's next move is a simple interstitial - interstitial hop is 3000 times the probability that it kicks out a Si; this is a factor of 4.6×10^6 reduction from the ratio if we had used 0.39 eV. Due to the reduction of the random walk migration of the Au_i, this change has the effect of expanding the depth scale for comparison with real experiments by a factor of 2.1×10^3; the time scale of the simulation is expanded similarly.

Of course, we must verify that the simulated sample is adequately thick to display all the physical phenomena that the input assumptions

imply. Thus, we have simulated both "thin" samples which are 1000 lattice constants, a, thick (i.e., 543 nm in the simulation but corresponding to 1.16 mm of real Si on account of the expansion factor) and "thick" samples which are 2000 a thick. We denote the depth direction as x. In the y and z directions we use cyclic boundary conditions, i.e., an object passing through the +y boundary reenters through the -y boundary with the same x and z coordinates and we allow defects near a boundary to sense defects on the other side. We set the y and z bounds to be + and - 100 a, but this has little significance. Thus, our thin samples have 6.4×10^8 lattice sites and our thick samples have 1.28×10^9 sites.

Thus, the "Q" for diffusion, the sum of the enthalpies of formation and migration, were taken to be 4.9 eV for Au_i and 4.8 eV for Si_i for the simulation of Ref. 40. The result turned out to be totally at odds with claims[47] of Seeger and others re: a) the Au_i profile which was not flat but exponentially decreasing into the bulk at all t; b) because the Au_i decays exponentially into Au-free bulk Si, it can not produce an asymmetric U shaped Au_{Si} profile with one-sided diffusion; c) the Au_{Si} profile was more exponential than the U shape claimed; and d) deep in the sample (in fact at all depths) $[Au_{Si}]$ was linear in t for all t and never showed the $t^{1/2}$ variation claimed.

We sought the advice of Prof. U. Gösele how the simulation parameters might be modified to obtain the behavior claimed in the KO literature. He opined[49] that we had assumed the Si_i to be too mobile and stated that it is necessary to achieve a large super saturation of Si_i in the bulk of the sample, which would occur if the Si_i are less mobile, in order to obtain the features claimed in the KO literature. Furthermore, he stated there is a critical ratio, R, of the diffusivities of Au_i and Si_i, $D(Au_i)$ and $D(Si_i)$, and their concentrations,

$$R = D(Au_i)\,[Au_i] \,/\, D(Si_i)\,[Si_i]\,, \qquad (3)$$

which must be larger than about 50 for the KO features to appear and that empirically seems to be in the range $100 < R < 200$. The thesis parameters given above imply $R = 2.3$.

Following Prof. Gösele's advice[49] we altered the assumed enthalpy of migration of Si_i to 0.94 eV, leaving the other parameters as above, which implies $R = 200$, and repeated the simulations. We now present those results. We first note that there is no qualitative difference between the present $R = 200$ results and the $R = 2.3$ results of the thesis; all the blatant discrepancies between the computer experiments and the properties claimed for the KO mechanism just cited remain. Moreover, using the simulation to observe the process in detail, one finds it easy to understand where and why these discrepancies arise. It seems to us that they are quite inevitable.

As our samples have cyclic boundary conditions for y and z and ideal Au_i and Si_i source/sinks on its x boundaries, it has no free surface where melting could nucleate. This means that our simulation cannot melt even though real Si in contact with a sufficient source of Au will melt to the Au-Si eutectic at T = 1368 K where $[Au_{Si}]$ reaches[48] about 1×10^{17} cm^{-3}. The KO literature does not take account of the melting of the Si during the diffusion experiment, which would introduce a moving boundary condition and greatly complicate their analysis and is assumed to be a small effect in the typical experiment.[49] Thus, we conclude it is proper for us to simulate the KO as we have done and allow $[Au_{Si}]$ to become unphysically large near the surface. (It would be a simple matter to change the sample specification periodically to eliminate surface regions that exceed the melting limit, but it is obvious that the effect of this would be to increase the discrepancy between the KO simulation and what has been claimed for it.)

After checking that our two-sided diffusion simulations are indeed symmetric to within the expected statistical accuracy, we conclude that

it is redundant to display both sides of the distribution. Instead we add the distributions symmetrically about the center in the x direction (x = 0) when presenting our results.

With this introduction in mind, we present in Fig. 1 $[Au_{Si}](t)$ for depths of 50 and 150 a for two-sided diffusion into the thin (1000 a) sample with R = 200 as suggested by Gösele. We note that: a) the variation is linear for all t; t extends to 2.4×10^{13} attempt periods or of order 2 sec.; and c) real Si would have melted through these 200 a long before the simulation stopped. In Fig. 2 we show $[Au_{Si}](t)$ for the same simulation at depths of 350 and 450 a. In order to check the result for the effect of statistical fluctuations, the simulation was repeated from the t = 1.15×10^{13} configuration on a different machine with different random number seeds. Those results are also plotted in Fig. 2 and compare as expected. We note that: a) the variation with t is linear, or if anything greater than linear and certainly not $t^{1/2}$, for all t; and b) the simulation continued to the point that real Si would have melted to a depth of 400 a from both sides, i.e., 80% of the sample would have melted. It is obvious that if we had moved the x boundary, which is the source of new Au_i, in accord with the melting that is implied, the rate of increase of $[Au_{Si}]$ in the remaining Si would be much faster than linear and not $t^{1/2}$.

In Fig. 3 we show $[Au_i](x)$ for various t for the thin sample diffusion. We note that the profile: a) is steeply sloped (approximately exponential) for all t even up to the point that the real sample would be 80% melted; and b) does not change much with t. As the same feature was observed in Ref. 40 (R = 2.3) with only quantitative differences, we conclude this feature is not a function of R. It simply results from the fact that if Au_i can displace Si_{Si} and it diffuses into relatively Au free regions, then it will do so and an exponential

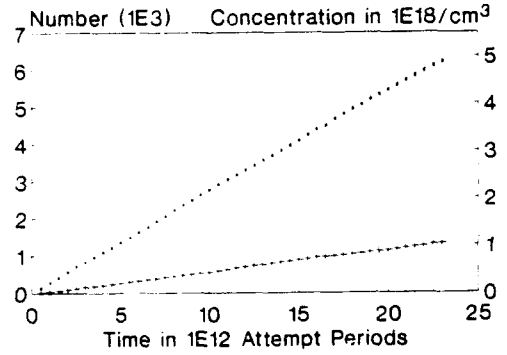


FIGURE 1
KO $[Au_{Si}](t)$ at depths of 50 and 150 a for two-sided diffusion into "thin" sample with R = 200 and spontaneous injection of Au_i.

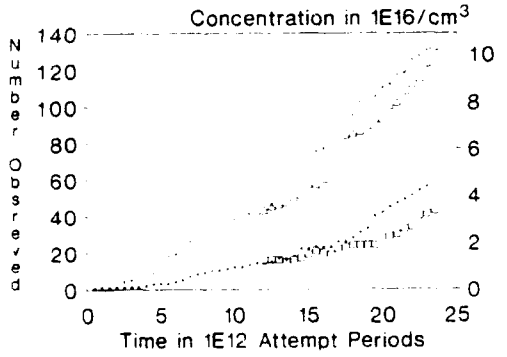


FIGURE 2
$[Au_{Si}](t)$ at depths of 350 and 450 a as in 1. Part of simulation is repeated to show statistical variation.

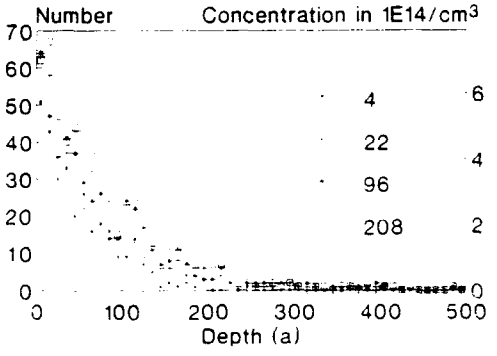


FIGURE 3
$[Au_i](x)$ for various t (1E11 attempt periods). To improve statistics 7 evenly and adequately spaced observations have been added.

decay profile will result. Contrary to what is asserted in the KO literature, the out diffusion of the Si_i that results from this displacement and any supersaturation of the Si_i has no qualitative effect on the processes. The exchange between Si_i and Au_i that happens many times for each Au_{Si} created deep in the sample merely renormalizes the effective diffusion coefficient of the Si_i. We regard this point as the most important fallacy of Ref. 47 and the literature derive therefrom. That literature assumes a flat profile for $[Au_i]$ is established almost immediately and maintained throughout the experiment.

In Figs. 4 and 5 we present the corresponding $[Au_{Si}]$ profiles for two-sided diffusion into a thin sample if we use SIM_KO instead of SIM_KI, i.e., assume the injection of new Au_i is stimulated by the arrival of Si_i at the gold source rather than being a spontaneous random event. To save time and because the shorter time simulation was compared in Ref. 40, we started from the $t = 1.15\times10^{13}$ configuration obtained from SIM_KI, as was used for Fig. 2. Here too we note that the variation of $[Au_{Si}]$ is linear for all depths and for all t to the point that the sample is 80% melted.

In Fig. 6a we compare the Au_i profiles for the thin and the thick sample for $t = 1\times10^{13}$. Actually, in order to increase the number of

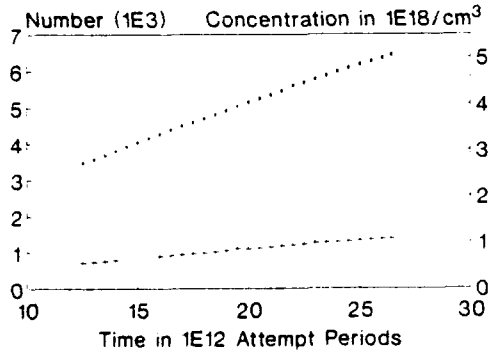


FIGURE 4
KO $[Au_{Si}](t)$ at x = 50 and 150 a for stimulated injection of new Au_i. All else the same as 1.

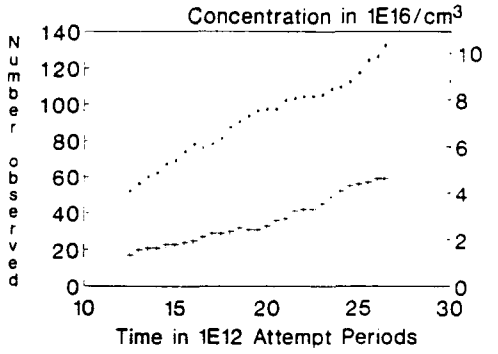


FIGURE 5
KO $[Au_{Si}](t)$ at x = 350 and 450 a for stimulated injection of new Au_i. All else the same as 2.

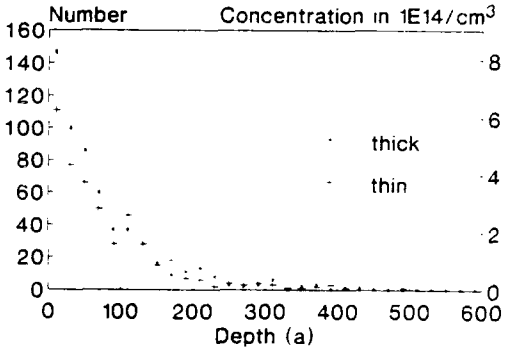


FIGURE 6a
$[Au_i](x)$ at t = 1E13 attempt periods for thick and thin samples. To improve statistics 7 evenly and adequately spaced observations have been added.

Au_i counted so as to improve the statitistics, we have added the values observed for seven observations spaced at t intervals of 5×10^{11} about this time. About 3×10^7 events, of which 95% are interstitial hops, occur in each interval so the Au_i distributions are thoroughly mixed between these intervals. We see that they are effectively the same and in fact we find that the total number of Au_i in the two samples is the same within statistical fluctuation. Thus, doubling the thickness of the sample has no effect on the number of Au_i

present under the KO assumptions. This is because effectively no Au_i's ever manage to penetrate more than 500 a, corresponding to 1 mm of real Si, from the eutectic interface. Thus again, the KO can not produce a U shaped profile from a one-sided diffusion. Figure 6b plots the same data with the depth scale extended for the thick sample to restate the point that no Au penetrates much beyond 500 a. Figure 7 plots $[Au_{Si}](x)$ for both thin and thick samples at $t = 1\times10^{13}$. Within statistical accuracy these are the same also.

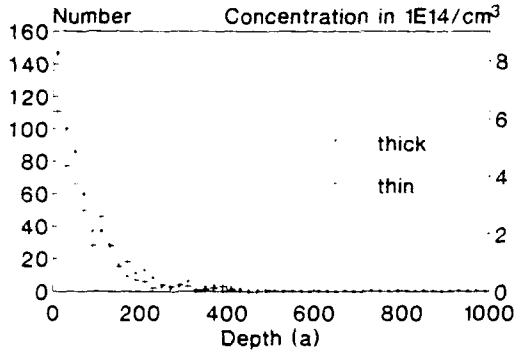


FIGURE 6b
Same as 6a but showing bulk of thick sample where no Au penetrates for KO assumptions.

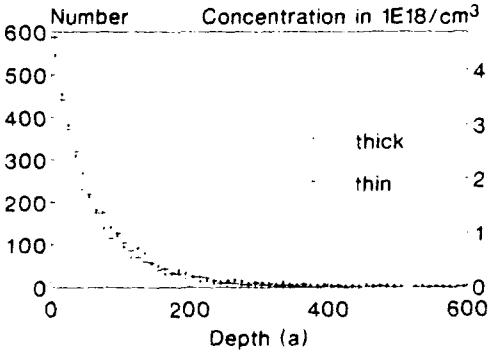


FIGURE 7
$[Au_{Si}](x)$ at t = 1E13 attempt periods for thick and thin samples.

3. CONCLUSIONS

Our computer experiments show that if we make the KO assumption that Au_i can displace host Si and produce Si_i, then when diffusing into bulk Si it will do so. This produces an exponential decay of $[Au_i](x)$ that is not affected by the thickness of the sample, by t, by choice of R, or by super-saturation of Si_i in the bulk of the sample. This contradicts a KO theory boundary condition, that $[Au_i]$ should be flat and near equilibrium with the eutectic for all t of interest. We are forced to conclude that all results of the large KO literature based on Ref. 47 must be completely revised. In particular, the fact that one sided diffusion of Au into Si does produce an asymmetric U profile for Au_{Si}, and the widely held assumption that Au diffuses mainly as Au_i, must imply that Au_i cannot produce Si_i to any appreciable extent even when diffusing through thick samples. The Au_i must get through the center of the sample and find some type of V_{Si} or vacancy complex that diffuses in from the Au free surface. This supports the atom-level information that Si_i has such a large enthalpy of formation that it plays no role in thermal diffusion processes.

REFERENCES

1 W. Kohn and L.J. Sham, Phys. Rev. 140 (1965) A1133.

2 O. Gunnarson, M. Jonson, and B.I. Lindqvist, Phys. Rev. B 20 (1979) 3136 and therein.

3 G.A. Baraff and M. Schlütter, Phys. Rev. B 30 (1984) 3460.

4 R. Car and M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.

5 M.T. Yin and M.L. Cohen, Phys. Rev. B 26 (1982) 3959.

6 R.M. Martin, Adv. Solid State Phys. XXV (1985) 3.

7 Y-T. Shen, D.M. Bylander and L. Kleinman, Phys. Rev. B 36 (1987) 3465 and therein.

8 W. Hanke, Th. Goelzer and H.J. Mattausch, Solid State Comm. 51 (1984) 23.

9 M.S. Hybertsen and S.G. Louie, Phys. Rev. B 34 (1986) 5390.

10 W. Hanke and L.J. Sham, Phys. Rev. B 21 (1980) 4656.

11 G. Strinati, H.J. Mattausch and W. Hanke, Phys. Rev. Lett. 45 (1980) 290.

12 W. Hanke, G. Strinati and H.J. Mattausch in Recent Developments in Condensed Matter Physics, edited by J.T. Devreese (Plenum, New York, 1981) Vol. 1.

13 J.A. Van Vechten and R.M. Martin, Phys. Rev. Lett. 28 (1972) 446.

14 W. Hanke and L.J. Sham, Phys. Rev. Lett. 33 (1974) 582.

15 S.G. Louie, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. Lett. 34 (1975) 155.

16 J. Dabrowski and M. Scheffler, Phys. Rev. Lett. 60 (1988) 2183.

17 D.J. Chadi and K.J. Chang, Phys. Rev. Lett. 61 (1988) 873.

18 S. Dannefaer and D. Kerr, J. Appl. Phys. 60 (1986) 1313.

19 S. Dannefaer, P. Mascher, and D. Kerr, J. Phys. Condens. Matter 1 (1989) 3213 and many therein.

20 M.J. Puska, O. Jepsen, O. Gunnarson, and R. M. Nieminen, Phys. Rev. B 34 (1986) 2695.

21 J.C. Bourgoin and J.W. Corbett, Radiat. Eff. 36 (1978) 157 and therein.

22 G.D. Watkins, in Deep Centers in Semiconductors ed. by S.T. Pantelides (Gordon Beach, New York, 1985) p. 147 and therein.

23 D.V. Lang in Deep Centers in Semiconductors ed. by S.T. Pantelides (Gordon Beach, New York, 1985) p. 489 and therein.

24 K. Bonde-Nielsen, H. Grunn, H. Haas, F.T. Pedersen, and G. Weyer, J. Electron. Mater. 14a (1985) 1065.

25 D.L. Williamson, J. Appl. Phys. 60 (1986) 3466 and therein.

26 J.A. Van Vechten, Phys. Rev. B 33 (1986) 2674.

27 J.A. Van Vechten in Handbook on Semiconductors, Vol. 3, edited by S.P. Keller (North Holland, Amsterdam, 1980) pp. 1 and therein.

28 D. Matiot and J.C. Pfister, J. Appl. Phys. 55 (1984) 3518.

29 J.A. Van Vechten, Phys. Rev. B 38 (1988) 9913.

30 M. Yoshida, Jpn. J. Appl. Phys. 27, 967 (1988).

31 P.M. Fahey, P.B. Griffin, and J.D. Plummer, Rev. Mod. Phys. 61 (1989) 289 and therein.

32 S.M. Hu, J. Appl. Phys. 57 (1985) 4527 and therein.

33 T.Y. Tan and U. Gösele, Appl. Phys. A 37 (1985) 1 and therein.

34 H.J. Lim, H.J. von Bardeleben, and J.C. Bourgoin, J. Appl. Phys. 62 (1987) 2738.

35 J.F. Wager and J.A. Van Vechten, Phys. Rev. B 39 (1989) 1967.

36 A. Seeger and K.P. Chik, Phys. Stat. Sol. 29 (1968) 455.

37 J.A. Van Vechten, J. Electron. Mater. 14a (1985) 293.

38 J.W. Matthews and J.A. Van Vechten, J. Crystal Growth 35 (1976) 343 and therein.

39 F.K LeGoues, R. Rosenberg, and B.S. Meyerson, Appl. Phys. Lett. 54 (1989) 751.

40 U. Schmid, Thesis, Department of Physics, Oregon State University, November 1988.

41 J.A. Van Vechten and U. Schmid, J. Vac. Sci. Technol. B 7 (1989) 827.

42 U. Schmid, N.C. Myers, and J.A. Van Vechten, Comp. Phys. Commun. in press.

43 F.C. Frank and D. Turnbull, Phys. Rev. 104 (1956) 617.

44 W.R. Wilcox and T.J. LaChapelle, J. Appl. Phys. 35 (1964) 240.

45 F.A. Huntley and A.F.W. Willoughby, Solid-State Electron. 13 (1970) 1231.

46 F.A. Huntley and A.F.W. Willoughby, J. Electrochem. Soc. 120 (1973) 414.

47 A. Seeger, Phys. Stat. Sol. A 61 (1980) 521.

48 E.R. Weber in: Properties of Silicon, EMIS Datareview (INSPEC IIE, London, 1988).

49 U. Gösele, private communication 1/20/89.

FAILURE OF THE "KICK-OUT" MODEL FOR THE DIFFUSION OF Au INTO Si WHEN TESTED BY MONTE CARLO SIMULATION

U. Schmid*, J. A. Van Vechten†, N. C. Myers†‡, and U. Koch†§

* Max-Planck-Institut für Festkörperforschung, Heisenbergstr. 1, D-7000 Stuttgart 80, Federal Republic of Germany.

† Center for Advanced Materials Research, Department of Electrical and Computer Engineering, Oregon State University, Corvallis, OR 97331-3211, USA.

ABSTRACT

We have performed large scale computer simulations on the controversial issue of Au diffusion into Si at T = 1095° C. Using a Monte Carlo algorithm and a conveniently parametrized set of parameters, our computer program is capable of working out the macroscopic consequences of a variety of models, i.e. atom level assumptions, in an unbiased way and without the approximations introduced in analytic calculations.

When applied to the "kick-out" hypothesis, our results are dramatically at odds with the properties claimed by its proponents. Neither the profile of the Au substitutionals, nor the Au-interstitial profiles are in agreement with the analytically obtained results. The discrepancy becomes most pronounced when comparing the variation with time of the Au concentration in the center of the sample, which we find to be linear at all times, in contrast to the alleged $t^{1/2}$ behavior. Moreover, the Au profile of a one-sided diffusion never becomes U-shaped, as experimentally observed.

1. INTRODUCTION

The understanding and control of diffusion in semiconductors is crucial to the progress in semiconductor electronics, and yet this is one of the least understood subjects in modern materials research. Major controversies rage about almost every aspect in this field, such as the role of self-interstitials and vacancies in atomic diffusion. For one part, this sad situation is due to the fact that atomic diffusion is a very complex subject, and thus the interpretation of experimental data can lead to contradictory reports. On the other hand, it can be attributed to the introduction of more and more models with an increasing number of parameters, which added confusion rather than clarification to the subject. Even the simpler models suffer from one major deficiency: the relationship between the atom level assumptions and its macroscopic consequences are extremely complex and not easy to establish. The connection between these two classes should be made in terms of thermodynamic reasoning. This task has been attempted in the case of vacancy hopping processes for Si and impurities by a few authors [1-5], but still all such work relies on intuition, contains approximations and passages to limits and is based on thermodynamic equilibrium assumptions for processes, which do not take place in equilibrium.

For this reason we conclude that only a computer program that takes into account the complex interaction of a diffusing particle with its vicinity can establish an exact and unbiased relation between microscopic assumptions and macroscopic consequences. Therefore we have developped the computer program VIDSIM (Vacancy and Interstitial Diffusion Simulator) [6, 7], that models the evolution of an initial set of particles in space and time for any of the chosen models the user wants to simulate. This is possible with the help of a "brute force" Monte Carlo algorithm, which involes an enormous amount of computational effort, if statistically significant results are to be obtained. Nonetheless, this effort can be achieved in a rather inexpensive way with the use of several ordinary PCs, which can work simultaneously on the same problem in the background while they are used for other purposes.

Mat. Res. Soc. Symp. Proc. Vol. 163. 1990 Materials Research Society

such as word processing. Of course, VIDSIM can also run on other computers, such as UNIX workstations and superminis.

We have designed VIDSIM in such a way, that virtually anyone who wants to test his model and parameters, such as activation energies, can use it, even if he or she is no computer expert. Its documentation is published in Ref. [8], and it is available from the Computer Physics Communications Library, catalogue number ABRF, or, for the cost of the medium, from the authors.

Monte Carlo simulations are not the only useful computational method for the investigation of diffusion processes in semiconductors. Some workers try to make progress with ab initio methods on supercomputers, with the help of calculations which are based on the local density approximation (LDA) [9]. While this method is well adapted to bulk ground state properties, it gives band gaps that are much smaller than the experimental excitation energies [10]. For this reason it is common practice to adjust the calculated band gaps to experiments by shifting the conductions bands with the so-called "scissors-operator", or by other "ad hoc" corrections [11, 12]. Whereas these corrections give extremely satisfactory results for band-structure calculations, it has been argued that the corrections needed for diffusion processes might be entirely different [13]. These calculations should be regarded with some care, as their uncertainty is typically more than 1.0 eV. In addition, such calculations can only be carried out by specialists on state of the art supercomputers, which makes them quite expensive and inaccessible to the experimentalist. This is especially true when the LDA is combined with molecular dynamics [14], which restricts calculations to just a couple of atoms during extremely short periods of time [15]. Some workers try to circumvent these restrictions by developping new theories and combine them with LDA calculations and available experimental data [16], or empirical tight-binding force models [15].

Such new techniques are very impressive and interesting and can be complementary to Monte Carlo methods, but with the problems mentioned above and the computer resources available today they have to be regarded as beyond the state of art. Practical progress is more likely to be made with semi-empirical theories.

2. VIDSIM ALGORITHM

VIDSIM performs a computer experiment on the theory which is to be tested. Our approach is simple brute force - the computer calculates the hopping probabilities for each mobile defect, based on the total free energy for the relevant process, which again depends on the assumptions and parameters the user inputs into the program. In addition to these parameters, the user has to supply the diffusion temperature T and an initial defect configuration, which can consist of both native point defects (vacancies and three different types of interstitials), and impurity related defects. The computer then adds up all partial probabilities of any process that can happen, and calculates the mean time for something to happen somewhere as the inverse of the total probability, in units of phonon cycles. According to the Mean Value Theorem, VIDSIM assumes that one diffusion process ("event") occurs after the calculated mean time, and chooses the event to occur with the help of a random number in the 0 to 1 interval from the set of possible events, which are weighted according to their normalized, individual hop probabilities. After the hop is completed, all possible events which are effected by this hop are reconsidered and the mean time for the next hop is re-evaluated. This process is repeated as many times as desired, or until a certain diffusion time is exceeded.

We prefer to run the same kind of simulation on various PCs and minicomputers simultaneously in order to obtain reliable statistics. We have also shown [13], that with this method even small university departments can obtain rates of computation which are equivalent to the speed of the order of one CRAY XMP or similar supercomputers all the time and as a single user! In fact, we have peformed a "multi Giga-event" simulation (several 10^9 discrete events) in a sample of 6.4×10^8 sites exclusively on PCs. This striking effect can be traced

down to the scalar nature of Monte Carlo algorithms, which do not allow vectorization or parallelization, and thus perform rather unfavorably on expensive vector-(super)computers.

An excellent example for the difficulties which arise when trying to deduce macroscopic properties from atom level assumptions is the diffusion of Au in Si.

3. THE DIFFUSION OF Au INTO Si CONTROVERSY

It has been known for a long time that, under thermal equilibrium conditions, Au is dissolved both on substitutional (Au_S) and interstitial sites (Au_i), and that during diffusion processes there is an exchange between them [17]. With the help of VIDSIM we now want to clarify the controversy about this mechanism. The first model, the "Frank-Turnbull mechanism" (FT) [18] involves vacancies (V) and gold interstitials (Au_i), which diffuse separately in the bulk and finally annihilate to form a substitutional site (Au_S):

$$Au_i + V \leftrightarrow Au_S. \tag{1}$$

Willoughby and co-workers concluded [19], that the FT was consistent with empirical observations that: a) the Au_S profile from a two-sided diffusion is more U-shaped than the result predicted by simple Fickian assumptions; b) the concentration of Au_S in the middle of this U-shaped profile, C_s^m, increases with time t roughly as $t^{1/2}$; and c) that an asymmetric U-shaped profile is obtained from one-sided diffusion.

The alternative model, the "kick-out mechanism" (KO), was proposed by Seeger and co-workers [20, 21] after analytical calculations, which seemed to indicate that the FT behaves more or less like a Fickian diffusion process, which is not in accordance with a) to c). The KO assumes that it is not the vacancies, but self-interstitials (Si_i), which determine the diffusion process. The Si_i is formed when an Au_i displaces a host Si atom:

$$Au_i + Si_S \leftrightarrow Au_S + Si_i, \tag{2}$$

and then migrates to the free surface, which is the only available sink in a highly perfect wafer. It is argued that it is the out-diffusion of Si_i that limits the increase of the central gold concentration C_s^m, and thus yields a $t^{1/2}$ behavior in agreement with b).

As the controversy about the two competing mechanisms has raged for many years, with different authors deriving contradictory conclusions from basically the same data, we decided to examine carefully the microscopic and macroscopic implications of the kick-out model without the approximations introduced in the analytic calculations [21], i. e. with a Monte Carlo computer experiment.

We also want to point out that neither the analytic calculations nor our computer simulation take into account any charge state effects of Si_i or Au_i, any Fermi level or recombination enhancements effects, and any elastic effects, and therefore still represents a great simplification of the real problem.

4. THE KICK-OUT SIMULATION

The simulation temperature was fixed at T = 1095° C, a value which is typically used in experiments and for which some combined models (a sum of FT and KO models) claim a predominant KO character of the diffusion process. For the injection conditions of the Au_i, we examined two extreme cases: a) the stimulated KO, in which an Au_i is injected into the surface at a random position every time an Au_S is formed in the bulk via reaction (2), this keeps the total number of Au_i in the sample constant; and b) the spontaneous case, in which the rate of injection occurs randomly, at a constant average rate. We have found no major qualitative difference between the two [6], and suppose that the true situation lies somewhere in the middle of these two. As it does not make much difference, we restrict ourselves to case b). We took the bond-centered interstitial for our simulation.

The necessary parameters for our simulation, enthalpies of formation and migration, were taken from the literature [22, 23], and we calculated the single bond enthalpy of Si and Au from their enthalpy of melting [2].

For practical reasons [6, 13], we increased the enthalpy of migration to 2.2 eV, compared to Weber's 0.39 eV. This is a concession to economy of simulation and examined in Ref. [6]. For the same reason, we had to choose the size of the simulated crystal smaller than what VIDSIM allows. We took the x-direction to be 1000 lattice constants, a, i. e. 543 nm, a value, which is expanded to about 1.2 mm by the increase of the enthalpy of migration [13]. To check that our sample is adequately thick, we repeated the simulation in "thick" samples, which are z = 2000 a. As we use cyclic boundary conditions for the other two directions, we can keep the y and z dimensions rather small (200 a).

Having examined our early results, which turned out to be totally at odds with claims by its proponents, we sought the advice of Prof. U. Gösele how he evaluates our choice of parameters. He argued [24] that our simulation did not achieve the supersaturation of Si_i necessary to reproduce the claimed behavior, as the Si_i is too mobile. He also stated that the critical ratio of the product of the diffusivity and concentration of Au versus that of Si, R, is much too low. Following his advice [24], we increased the enthalpy of migration of Si_i from 0.39 eV to 0.94 eV. Thus, our simulation parameters are now in accordance with the recommendations of the KO advocates.

5. RESULTS

Our earlier results on that issue have been presented elsewhere [13]. They have clearly shown that:

- the profile of the Au_i is steeply sloped (approximately exponential) for all t even up to the point that the sample would be basically melted. This feature is not affected by t, nor by choice of R, nor the thickness of the sample, and contradicts a KO boundary condition, that $[Au_i]$ should be flat and near equilibrium with the eutectic layer for all t.
- the variation with time of the Au_{Si} concentration in the center of the sample, C^s_m, is linear for all t, which extends the order of 2 sec., and never becomes $t^{1/2}$.
- the Au_{Si} profile itself, which we find to be much more exponential and less U-shaped than claimed.

These results have been obtained under the assumption that there is except for reaction (2) no annihilation process (sink) for the Au_i, so that the concentration of Au_{Si} grows steadily on the surface. This results in the progressive melting of the crystal from the surface on. We note that none of the above results change qualitatively, if we assume that the Au_i annihilate at the surface, so that the surface acts as a sink for both the Si_i and Au_i.

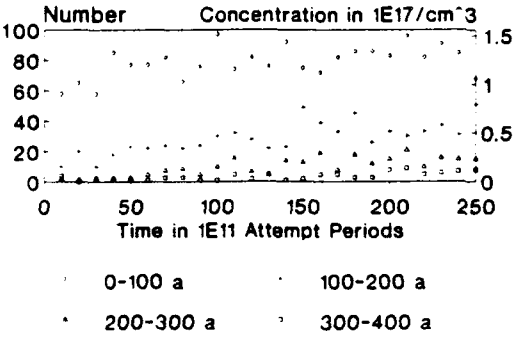


Fig. 1 One sided diffusion: [Au$_i$] versus time for 4 different layers. The injection takes place at x = 0 a. At greater depths all numbers are zero.

Experiment shows, that a one-sided diffusion of Au into Si produces an asymmetric U-shaped profile. The simulation of the kick-out mechanism for a source on one side only with annihilation of all interstitials leads to the results and Figs. 1 - 3.

In Fig. 1 we present the time evolution of the number of Au_i for 4 different layers, each 100 a thick. The layer which is closest to the surface (0-100 a) contains the

highest concentration of Au_i, of course. We see from this plot that the number of interstitials per layer increases perceptively. Initially, the $[Au_i]$ is determined by the balance between injection of Au_i from the source and their annihilation at the surfaces or displacement of Si. As time increases, Au_i are also created by the back reaction (2). We see that the total number of Au_i has increased a factor of two in the course of the simulation.

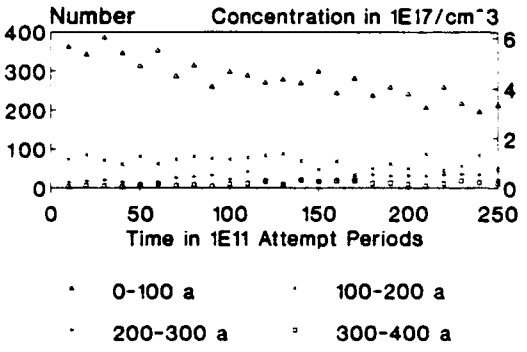


Fig. 2: One sided diffusion: $\Delta[Aus_i]$ (t) for 4 different layers.

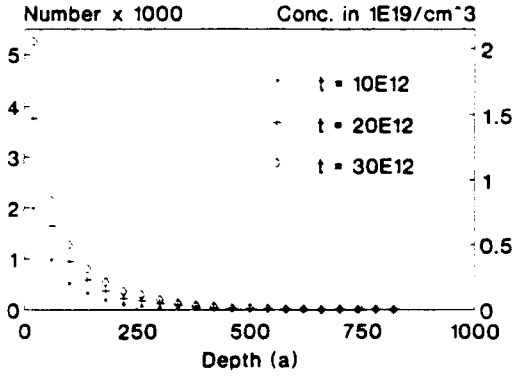


Fig. 3: $[Aus_i]$ from a one-sided diffusion for different times.

In Fig. 2 we depict the increment of $[Aus_i]$ for various layers. For the surface region (0-100 a) one can cleary see the saturation of $[Aus_i]$. The two-sided diffusion again failed to find any evidence for the $t^{1/2}$ behavior of the central gold concentration.

Fig. 3 shows a Aus_i profile for 4 different diffusion times, in units of attempt periods. The diffusion started from the left side at x = 0 a. Our simulation does not reproduce the experimental U-shaped profile. Although further investigation has to be done on that issue (the simulation has to run a little longer), we are convinced that the basic features of this profile will not change, when the simulation is based on the KO assumptions, no matter how parameters are varied. We think the rise in $[Aus_i]$ at the far side of the sample can only occur if Au_i pass through the center of the sample without creating Si_i, to find V coming from the far side. This occurs only if $\Delta H^f(Si_i)$ is too high for Si_i to play a role.

6. CONCLUSIONS

We have performed a multi Giga-event simulation of the kick-out theory on PCs. Even with the parameters as suggested by the proponents of that theory, we could not reproduce any of the features claimed by them. We have shown that the exponential decay of the Au_i profile is neither affected by the thickness of the sample, nor by t, nor R. Based on our latest results, we think the fact that the surface acts as a sink for the Au_i has little effect on the simulation. One major problem of the KO theory and the parameters we tested can be traced down to the high probability with which an Au_i can displace a Si. It's "scattering cross section" is simply too high. Thus it can not penetrate deeply into the sample, and its profile decays exponentially. Another major point we have eludicated is the one-sided diffusion issue. It is impossible to obtain an asymmetric U-shaped profile with the KO hypothesis.

Based on our results, we think it is appropriate to exclude the KO as one of the possible mechanisms which mediate the Au in Si diffusion.

ACKNOWLEDGEMENTS

Supported in part by A.F.O.S.R. under contract No. 89-0309. Presentation at this conference made possible by "Stifterverband für die Deutsche Wissenschaft", for which U. S. is especially grateful. Two of us (U.S. and U.K.) thank the Fulbright commission and the Land Baden-Württemberg for a scholarship. Valuable discussions with U. Gösele (Duke University), P. Reineker and H. Jex (Universität Ulm) are acknowledged.

REFERENCES

‡ Present address: Mentor Graphics, Inc., Beaverton, OR, USA.
§ Present address: Institut für Informationsverarbeitung der Universität Tübingen, D-7400 Tübingen, FRG.
[1] J.A. Van Vechten, Phys. Rev. B **33**, 2674 (1986).
[2] J.A. Van Vechten, in *Handbook on Semiconductors*, Vol. 3, edited by S. P. Keller (North Holland, Amsterdam, 1980), pp. 1 and therein.
[3] D. Mathiot and J. C. Pfister, J. Appl. Phys. **55**, 3518 (1984).
[4] J. A. Van Vechten, Phys. Rev. B **38**, 9913 (1988).
[5] M. Yoshida, Jpn. J. Appl. Phys. **27**, 967 (1988).
[6] U. Schmid, Thesis, Dept. of Physics, Oregon State University (1988); Diplomarbeit, Abt. Theoretische Physik, Universität Ulm (1989).
[7] J. A. Van Vechten and U. Schmid, J. Vac. Sci. Technol B **7**, 827 (1989).
[8] U. Schmid, N. C. Myers, and J. A. Van Vechten, Comp. Phys. Commun., in press.
[9] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[10] G. B. Bachelet and N. E. Christensen, Phys. Rev. B **31**, 879 (1985).
[11] N. E. Christensen, Phys. Rev. B **30**, 5753 (1984).
[12] U. Schmid, N. E. Christensen, and M. Cardona, Phys. Rev. B submitted.
[13] J. A. Van Vechten, U. Schmid, and N. C. Myers, in *Proceedings of the International Conference on the Science and Technology of Defect Control in Semiconductors*, edited by K. Sumino (North Holland, 1989), and therein.
[14] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[15] C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. B **39**, 8586 (1989).
[16] C. S. Nichols, C. G. Van de Walle, and S. T. Pantelides, Phys. Rev. Lett. **62**, 1049 (1989); Phys. Rev. B **40**, 5484 (1989).
[17] W. R. Wilcox and T. J. LaChapelle, J. Appl. Phys. **35**, 240 (1964).
[18] F. C. Frank and D. Turnbull, Phys. Rev. **104**, 617 (1956).
[19] F. A. Huntley and A. F. W. Willoughby, J. Electrochem. Soc. **120**, 414 (1973); Solid-State Electron. **13**, 1231 (1970).
[20] A. Seeger, Phys. Stat. Solidi (a) **61**, 521 (1980).
[21] U. Gösele, W. Frank, A. Seeger, Appl. Phys. **23**, 361 (1980).
[22] T. Y. Tan and U. Gösele, Appl. Phys. A **37**, 1 (1985).
[23] E. R. Weber in: *Properties of Silicon*, EMIS Datareview (INSPEC, IIE London, 1988).
[24] U. Gösele, private communication, Jan. 1989.

Journal of Electronic Materials, Vol. 20, No. 5, 1991

Surface Treatment Effects on Atomic Diffusion in Si Explained Without Self Interstitials*

J. A. VAN VECHTEN, U. SCHMID** and ZHANG Q.-S.

Center for Advanced Materials Research
Electrical and Computer Engineering Dept.
Oregon State University, Corvallis, Oregon 97331

Many recent discussions of surface treatment effects on atomic diffusion in Si have explained these largely in terms of effects attributed to Si self interstitials. However, we have shown by our straight forward Monte Carlo (VIDSIM) simulation of diffusion of Au into Si according to the "kick-out" mechanism of Seeger that this mechanism is in fact completely incapable of explaining the two-sided, "U shaped" profile of substitutional Au which results from a one sided in-diffusion of Au. We have shown that if Au interstitials can displace Si at any appreciable rate, then the Au substitutional profile must decrease monotonically from source side to far side. We noted that this is strong evidence that Si self interstitials play no role in thermal processes in Si. Here we show that the surface treatment effects often attributed to Si self interstitials can be naturally explained without them.

Key words: Si, atomic diffusion, surface treatment effects

I. INTRODUCTION, Au DIFFUSION INTO Si

There has long been a controversy re the relative importance of vacancies, V, or self interstitials, Si_i, to thermal process in Si.[1-14] Those who concentrate on studies at the atomic level tend to believe that Si_i has too large a heat of formation, and thus too small an equilibrium concentration, to play any role.[3,4,8,9] They note that V has been unambiguously observed by magnetic resonance[8] and by positron methods[15] after the creation at low temperatures of $V + Si_i$ pairs by electron irradiation, but that the Si_i, which must have been created at the same time, has never been seen. Instead, resonance methods[8] find impurity interstitials, *e.g.* B_i or C_i, created in numbers equal to that for V. They also observe that:[4] a) 12 eV must be transferred from the bombarding electron to the Si atom on a lattice site to create the pair[16] b) convincing theoretical arguments,[17,18] which account for the variation of the rate of displacement with the angle between the beam and the crystal axes, show that the excess energy of this irreversible process over the thermodynamic heat of formation for the pair can not be more than 1 eV, leaving 11 eV for the sum of the heats of formation of V and Si_i, $\Delta H_f(V) + \Delta H_f(Si_i)$; and c) substantial concentrations ($>10^{16}$ cm^{-3}) of V are seen in thermal equilibrium in Si at temperatures well below the melting point by both positron annihilation[19] and "Simmons-Ballufi" methods[20,21] (comparison of thermal expansion of the lattice constant with thermal expansion of the sample volume to determine if lattice sites are being added or subtracted).

A comment should be made re point c above. While there is debate over the magnitude and temperature dependence of the trapping cross sections, σ, of the (ionized) V's in Si,[19,22] which directly affects estimates of the V concentration, $[V]$, there is no question of calibration for the Simmons-Ballufi experiment; recent results[20] show $[V] > 10^{16}$ cm^{-3} already at $T = 1300$ K. This contradicts the assumption made by Dannefaer *et al.* in analyzing positron data[19] that the Lax giant trapping cross section criterion, with its $\sigma \propto T^{-2}$ temperature variation, should be applied to $\sigma(V)$ at these T's. Thus, Dannefaer *et al.*'s estimated $\Delta H_f(V) = 3.6$ eV is too high by about 1.0 eV[22] just as their estimated $[V]$ is too low by a factor of 10^2. Even if $\Delta H_f(V)$ were as much as 4.0 eV, rather than the current best estimate[9,22] of about 2.7 eV, we would still have $\Delta H_f(Si_i) = 11 - 4$ eV $= 7$ eV, which would mean Si_i would play no role in thermal processes. Recall that activation energies for self diffusion in Si range with T[23] from 4.1 to 5.2 eV. (The activation energy for Si_i migration must be finite and positive. The term which Si_i contributes to the total Si self diffusion has an activation energy that is the sum of this positive migration energy plus the energy of formation. Even if one were to suppose the migration energy for Si_i were almost zero so that Si_i migrates out of the sample almost instantly after creation with a formation energy of 7 eV, they could not produce a Si self diffusivity with an activation energy of 5 eV, nor even contribute significantly to the self diffusion.)

Such considerations lead to the suggestion that, if any additional point defect beyond the V (with its five ionization states and complicated behavior) is required to understand thermal processes, in Si, then

*Supported in part by U.S. AFOSR-89-0309.
**Present address: Max Planck Institut für Festkörperforschung, Stuttgart, Germany.
(Received November 21, 1990; revised January 15, 1991)

0361-5235/1991/1401-43135 00©91 TMS

these may be "extended interstitials"[1] or "amorphous zones"[9] rather than the Si, which must exist for an instant too short to be observed in irradiation experiments. The entropy of Si self diffusion is observed to be 6 to 10 k at high temperatures. Such large values have been interpreted as evidence either for extended defects (amorphous zones)[1] or for an entropy of ionization of the V comparable to that of the band gap.[9]

On the other hand, many of those who concentrate on studies at the device scale have invoked an unspecified type of "Si self interstitial," which might or might not be an "extended" or "amorphous zone", and ascribe adjustable parameters to it in order to describe data and phenomena of interest.[5-7,11,12] They sometimes claim that the V can not play the majority role in many thermal processes in Si.

Much of the argument has revolved around the diffusion of Au into Si, which is widely studied[24-26] because of the dramatic effect Au_{Si} has on minority carrier lifetimes. The following points re this subject are not controversial: a) Au diffuses very rapidly into Si, and into Ge and other tetrahedral semiconductors, as an interstitial, Au_i. (Although Au_i has not been definitely observed or characterized, no one debates its role.) b) The dominant site for Au atoms is substitutional, Au_{Si}, where it acts both as a deep donor and as a deep acceptor. c) Even when Au is diffused in strictly from one side only, the resulting Au_{Si} concentration profile, $[Au_{Si}]\ (x)$ is not at all monotonic but essentially symmetric with a broad flat minimum in the center of the sample. $[Au_{Si}]\ (x)$ is said to be "U shaped." See Fig. 1. Any remaining doubt that this U shaped profile which results from what was intended to be a one sided in diffusion might be an artifact of undetected surface diffusion has been erased by ion implantation experiments (with capping layers and facing wafers) reported by Coffa *et al.*.[27] (Si_3N_4 and SiO_2 capping layers prevented the ion implanted Au from getting to the near surface; this was verified by showing that no Au reached a polished bare wafer in intimate contact with the near surface during the diffusion.) d) $[Au_{Si}]$ at the center of the sample increases roughly as the square root of diffusion time, t, for values of t intermediate between that required for Au_{Si} first to appear and that required to reach half saturation. e) The diffusion of Au into Ge from one side also produces a two sided profile, $[Au_{Ge}]\ (x)$, but the shape is noticeably different from $[Au_{Si}]\ (x)$. f) Several other transition metals produce substitutional profiles similar to $[Au_{Si}]$ upon diffusion into Si. Re point e) it should be noted that diffusion processes in Ge have been much more clearly established and are generally agreed to be consistent with a pure vacancy model. This is largely because Ge isotopes make radiotracer experiments much easier for Ge than for Si.[9,23] Also the vacancy in Ge does not exhibit the negative-U behavior that Si vacancies do.[8] Thus, one can argue that the distinctions between the Au profiles into Si and into Ge result from difference in the properties of the vacancies in the two host or differences in the identity of the major defect.

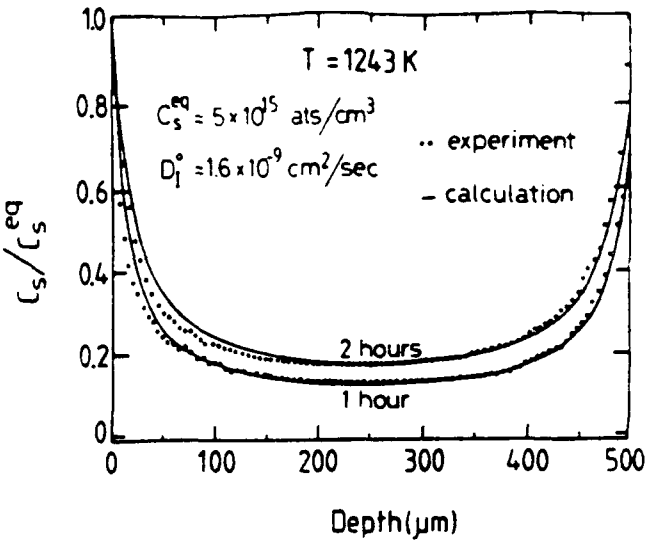


Fig 1a — Au substitutional profile normalized to the saturation value in Si reported by Coffa *et al.*, Ref. 27, from an ion implantation experiment in which the possibility of surface diffusion contributing to the peak near the far surface was carefully excluded. The profile is very similar to those obtained when a Au-Si molten eutectic is formed on one side. This is remarkable because the ion implantation injects a fixed number of atoms while variable fractions of the eutectic may be dissolved; also the boundary condition would seem to be distinctly different. The "calculation" plotted is that asserted by Gösele *et al.*, Ref. 6, for the boundary condition of constant infinite Au concentration on both surfaces. The asserted formula can be expressed as erf (ln(y)) = bx, where erf is the error function, b is a constant and y is the square root of $C \equiv [Au_{Si}]$. *See text.*

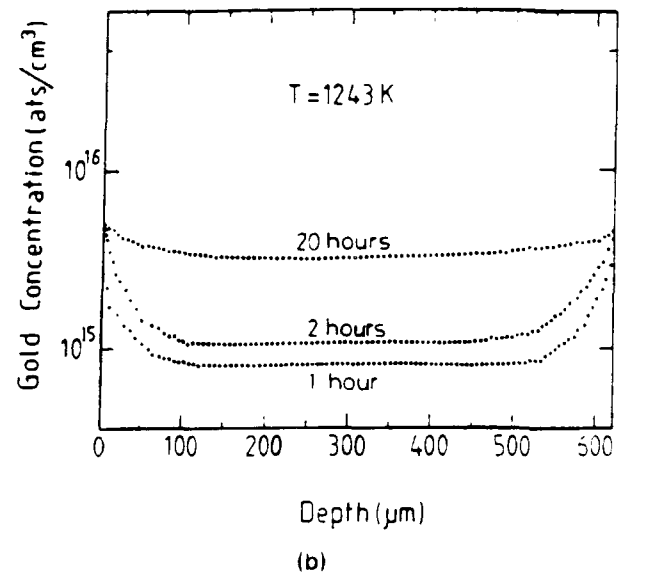


Fig. 1b — More data from same experiment of Coffa *et al.*, Ref. 27. Note that here a log scale is used where 1a used a linear scale.

The original interpretation[25,26] of these facts was that the "Frank Turnbull Mechanism,"[28,29]

$$Au_i + V \Leftrightarrow Au_{Si}, \qquad (1)$$

was the dominant mode of diffusion of Au into Si as well as for Au, Cu, etc. diffusion into Ge, for which

the mechanism was originally proposed and is generally accepted. Those who press this interpretation explain the two sided nature of $[Au_{Si}]$ (x) from a strictly one sided source by observing that V can be created at either the source surface or the far side surface and that it is widely agreed that Au_i diffuses very fast. Thus, the rate of production of Au_{Si} is limited by the in diffusion of V. $[Au_i]$ is assumed to be effectively uniform through the sample for all but very short t. The differences between the details of the shape of the profiles in Si versus those in Ge are attributed to the differences in the detailed behavior of V_{Si} and V_{Ge}. V_{Si} exhibits "negative U" behavior of its donor ionization states, five rather than four ionization states, more complex multivacancy behavior, and recombination enhanced diffusion.[8,30] However, to our knowledge no one previously has claimed to show how the difference in the detailed behavior of the V's accurately account for the differences in the details of the profiles.

On the other side, W. Frank, Gösele, Seeger and others have claimed[5,6] that (1) absolutely can not account for the distinctions between the profiles in Si and in Ge. They argue that instead a "kick out mechanism,"

$$Au_i + Si_{Si} \Leftrightarrow Au_{Si} + Si_i, \qquad (2)$$

must dominate the production of Au_{Si} and that the rate of the net process is limited by the out diffusion of Si_i, rather than the in diffusion of V_{Si}. If this were so, it would follow that the Si_i would have to be important to other thermal processes as well and that $\Delta H_f(Si_i) \ll 7$ eV.

Those who press (2) as the dominant process leading to Au_{Si} production have claimed to account for $[Au_{Si}]$ (x), and other phenomena, accurately in terms of this process. However, when we directly simulated their model with a completely straight forward Monte Carlo simulation procedure, VIDSIM,[31] [31], we found that the true out come of the model for a one sided diffusion is completely at odds with the claims of its advocates, and with empirical $[Au_{Si}]$ (x). We tried several variants of assumptions of model parameters and boundary conditions and found that no such variation could change the conclusion just stated.[13,14] Instead, we have identified fundamental inconsistencies in the arguments used to argue a connection between the kick out model and equations that do in fact approximate the empirical profiles. See Fig. 1a.

We now will briefly describe and clarify the errors in that reasoning. We focus the discussion to the case that the Au source is on one side of the sample only because this gives the clearest and most dramatic distinctions between the true result of the kick out model and what has been claimed for it.

Since the ion implantation experiments of Coffa *et al.*[27] have excluded all possibility of surface diffusion accounting for the appearance of Au_{Si} near the far surface, Au must have diffused through the bulk of the sample, presumably as Au_i. First let us suppose that the Au_i got from the source side to the far side without stopping to form Au_{Si} on the way. (This is the assumption of the Frank Turnbull alternative.) For some point in the center of the sample at a time early in the process, when very little Au_{Si} has been produced, consider the relative probability that a particular Au_i will hop one step further as an interstitial rather than displace a host atom near it, which are almost all Si. This probability must be very small or the Au_i will not make it to the far side of the sample before a displacement event occurs and the Au atom ceases to be an interstitial. Furthermore, this probability is directly linked to heats of formation of Si_i, Au_i and Au_{Si}. Let us denote the mean number of times Au_i hops as Au_i before it displaces a Si from the lattice as N^h. To make a numerical estimate, let us suppose stable site for Au_i is the bond centered interstitial site. (This choice will have little quantitative effect on the argument.) Then the distance of each Au_i-Au_i hop is $\sqrt{2}\ a/4$ or 0.192 nm for Si. These hops are assumed to be a random walk. The samples used by Coffa *et al.* were about 0.5 mm thick; others used much thicker samples and got the same result, but with perhaps less confidence that surface diffusion had been excluded. A well known result for a random walk is

$$\langle z^2 \rangle = 2\,D\,t = 2/3\,I^2\,N^d, \qquad (3)$$

where z is the displacement in one direction which we here take to be 0.5 mm, D is the diffusivity, I is the step length, which we here take to be 0.192 nm and N^d is the number of steps, which must at least equal the mean number of steps, before displacement, $N^d \geq N^h$. Thus we have

$$N^d = 1.5\,\langle z^2 \rangle / I^2 = 1.02 \times 10^{13} \qquad (4)$$

and $\ln(N^d) = 30$. The sum of the heat of formation of Si_i plus that of Au_{Si} must exceed that of Au_i by 30 kT for this assumption, that Au_i diffuses through the sample without displacing a Si from the host lattice, to be true. Let us take the diffusion temperature T = 1368 K, a typical value. Then 30 kT = 3.54 eV. The fact that $[Au_{Si}]$ is observed in the mid 10^{17} cm^{-3} range while Au_i is not directly observed at all implies that $\Delta H_f(Au_i) - \Delta H_f(Au_{Si})$ must be significant. The data analysis of Weber[32] estimates the value of this difference to be 2.5 eV. If this is so, then we would have $\Delta H_f(Si_i) > 3.5 + 2.5 = 6.0$ eV, which is in accord with the argument of those who advocate the Frank Turnbull mechanism and so large that the Si_i would play no role in thermal process.

Indeed, those who advocate the kick out mechanism assume that, contrary to the assumption of the last paragraph, the Au_i displaces Si many times before it diffuses from one side of the sample to the other. Thus, any Au_i that appears near the far surface to produce Au_{Si} there must itself already have been displaced from lattice sites in the bulk of the sample several times. The kick out hypothesis is that this occurs according to reaction (2) and via the action of the Si_i's.

This brings us to the question why does $[Au_{Si}](x)$ rise near the far surface? According to those who advocate the kick out hypothesis, this is because $[Si_i]$ is equal to its equilibrium value near either surface but is highly supersaturated in the middle of the sample and, further, occurs because the Si_i do not diffuse away as fast as the Au_i, which create them by reaction.[2] (The implication that the migration enthalpy, $\Delta H_m(Si_i)$, for Si_i must be significant reduces the maximal value for $\Delta H_f(Si_i)$ for which Si_i could play a role in thermal processes even further. Thus, if $\Delta H_m(Si_i) > 1$ eV, then $\Delta H_f(Si_i)$ must be less than 4 eV for Si_i to contribute to Si self diffusion with a component of activation enthalpy 5 eV.) It is precisely this assertion by those who advocate the kick out model that we have tested by Monte Carlo simulation and have shown to be incorrect.

The central error introduced in the reasoning of those who advocate the kick out model comes when they try to apply conditions that would apply once equilibrium has been reached to analysis of this system that, like all systems where there is net diffusion, is far from equilibrium. This appears in two related forms. One is in the assertion of what those authors call "local equilibrium" and means that the reverse rate of reaction (2), R_R, *i.e.* of

$$Si_i + Au_{Si} \Rightarrow Au_i \qquad (2r)$$

is essentially the same as the forward rate, R_F, *i.e.* of

$$Au_i + Si_{Si} \Rightarrow Au_{Si} + Si_i \qquad (2f)$$

at all points and at all times after the very beginning of the process. That is they assume as a boundary condition for their analysis that

$$R_R(x,t) \sim R_F(x,t) \text{ for all } x \text{ and } t. \qquad (5)$$

The second false assertion, which is really not independent of the false assertion of "local equilibrium," is that $[Au_i](x)$ is essentially flat and near its equilibrium value across the sample, *i.e.*

$$[Au_i](x) \sim [Au_i]^e \text{ for all } x \text{ and } t. \qquad (6)$$

This too is assumed as an initial condition in their analysis. It is easy to see that a consequence of (5) and (6) is that, if $[Au_{Si}](x)$ is U shaped, then $[Si_i](x)$ must be an inverted U of the same shape and magnitude.

When we do the straight forward Monte Carlo simulation of the kick out model, the result depends upon details of the surface boundary conditions, as one should expect and as we shall discuss below. However, for no set of surface boundary conditions does the simulation find the ASSUMPTIONS (5) and (6) invoked by the advocates of the model to hold.

Instead, we invariably find that when Au_i diffuse into a region that is relatively pure Si, then $[Au_i](x)$ decays monotonically and approximately exponentially. We feel this result is quite obviously the consequence of the assumption that Au_i can displace Si. If Au_i can transform to $Si_i + Au_{Si}$, then it does. We suggest that there is an analogy with the absorption of visible light; although the photons go as fast as anything can, the fact that they have an absorption length that is small compared with the width of the sample inevitably implies that they will decay exponentially in the sample and that very few will reach the far side.

The advocates of the kick out model opine that the reverse reaction (2r) (the analog of radiative recombination for the case of light), will occur at essentially the same rate, assumption (5), and will keep $[Au_i](x)$ flat and constant. We find that (5) is obeyed only when and where $[Au_{Si}]$ has reached equilibrium with the source. This occurs arbitrarily early in the process for points close enough to the source side surface, but does not occur at greater depths until $[Au_{Si}]$ has saturated. Again, we feel that this empirical finding is obvious and inevitable.

In order to obtain their ASSUMPTION (5), the advocates have to assume that $[Si_i]$ is very big where $[Au_{Si}]$ is very small, *i.e.* in the center of the sample for times short compared with that required to saturate the whole sample, which is very much longer than that required to saturate the far side surface region. We do not find any significant supersaturation of Si_i in Au_{Si} free regions for any plausible choice of surface boundary conditions. We note that when Si_i is diffusing in pure and perfect Si, *i.e.* where $[Au_{Si}] = 0 = [V]$, the diffusion must be conservative;

$$\frac{d^2}{dx^2}[Si_i](x) = 0. \qquad (7)$$

It is well known that (7) admits only linear solutions,

$$[Si_i](x) = a + bx. \qquad (8)$$

Furthermore, the value of $[Si_i]$ at the far surface, where $[Au_{Si}] = 0$ initially so that (7) will apply until some Au_i manages to arrive despite its exponential decay, is most likely at its equilibrium value for all t. The advocates assume this and also that it maintains the equilibrium value at the source surface as well. Thus, the inverted U profile for $[Si_i](x)$, which is required for the assumptions of the advocates, can not occur.

To summarize what does in fact happen with the kick out model when $[V] = 0$ everywhere and the source is on one side is the following. $[Au_i](x)$ at first decays essentially exponentially into the Au_{Si} free bulk. The Si_i produced by the displacement may either diffuse deeper into the Au free bulk, in which case their concentration is conserved and they simply diffuse until they are annihilated on the far side, or they diffuse back toward the source side, in which case they may either produce the reverse reaction and promote further diffusion of the Au or they annihilate at the source side surface. The resulting $[Au_{Si}](x)$ decreases monotonically from the source to

the far side for all times. Equilibrium between the forward and reverse reaction occurs first at the source side surface and is attained at monotonically increasing times for increasing depth. Equilibrium or reaction (2) is last attained at the far side surface. See Fig. 2.

In addition to its failure to produce a two sided profile from a one sided source, the kick out model, when straight forwardly simulated by Monte Carlo means, also fails to get the shape of the profile near the source side surface right. This can be illustrated for either a one side or for a two sided source. See Fig. 3. This finding is again contrary to the claims of those who advocate the kick out model because they errored when they assumed that local equilibrium, (5), would occur in the center of the sample *essentially as soon as it occurs in the source surfaces.*

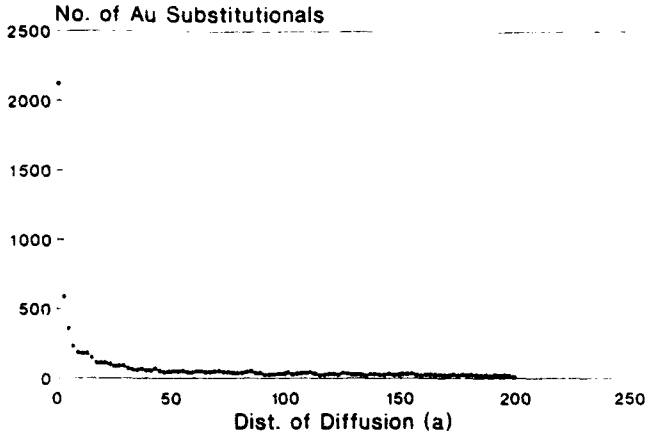


Fig. 2a — Result of direct simulation of the kick-out process, with no vacancies permitted, under the boundary conditions that a Au source on one side ($x = 0$) injects Au_i into the crystal randomly and at a constant rate and that any interstitial passing through either surface is annihilated. The reciprocal boundary condition of the generic VIDSIM simulator has been turned off so that Au_{Si} formed on the $x = 0$ boundary are not transferred to the far side surface. Here results are summed from three samples each 10 lattice constants, a, across in the y and z directions and 200 a in the x direction. As this computer simulation does not allow the crystal lattice to melt, the eutectic, that in fact forms as any constant Au source dissolves the surface of the Si, appears as a layer on the source side with $[Au_{Si}] > 10\%$ and a rather abrupt edge. At this point in the diffusion process that edge is at $x = 2$ a. As we let the process age beyond this point, we observe the eutectic layer become thicker. The slope of the profile in the region beyond the eutectic gradually decreases but remains monotonic.

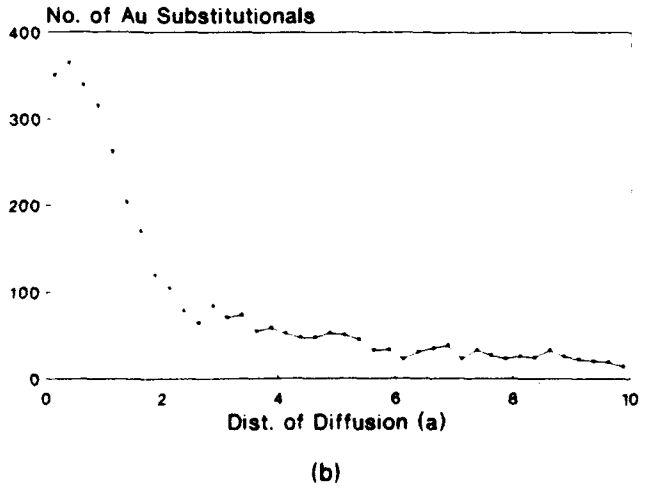


Fig. 2b — Near surface, expanded scale date from Fig. 2a.

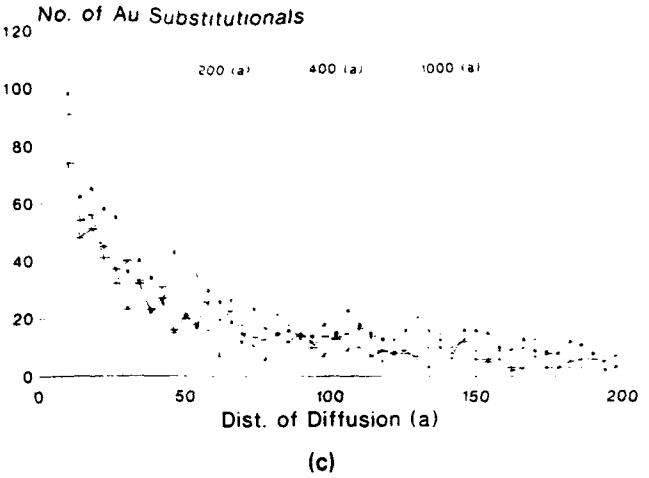


Fig. 2c — Results as in Fig 2a for the first 200 a depth of sample 200, 400 and 1000 a thick. In the thicker samples the $[Au_{Si}]$ profiles simply continue to die off monotonically with depth. Note the near source values of $[Au_{Si}](x)$ decrease with increasing sample thickness. This is because the closer the far side surface that annihilates the Si_i's, the less likely is reaction (2r) to displace a given Au_{Si}.

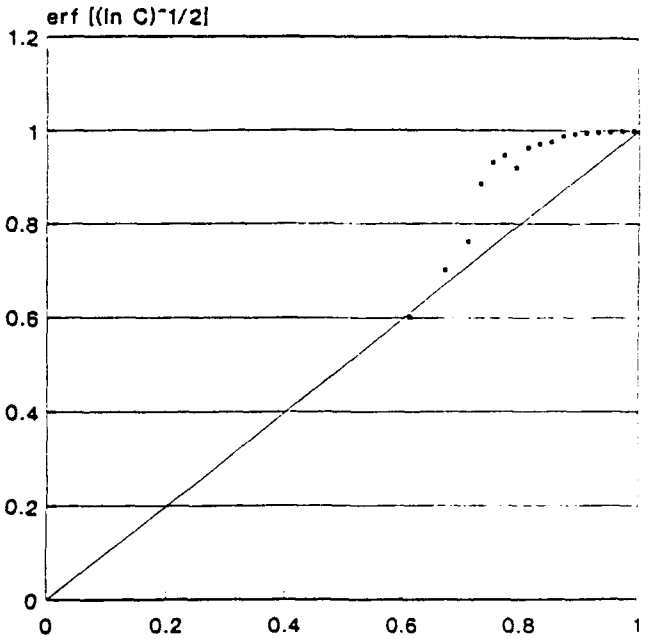


Fig. 3 — Result of direct simulation of kick-out mechanism under conditions assumed by Gösele *et al.*, Ref. 6, as reported in Ref. 33, dots showing stark contrast to the asserted result, the straight line. The sample is 1000 a thick with sources at both surfaces, $x = 1.0$. The center of the sample is at $x = 0$. Results from the two sides are statistically equivalent and have been added for display.

Because the use of a computer to simulate these processes releases us from the constraints of physical reality, we can manage a (completely artificial) situation in which the assumption (5) is reasonably well satisfied near one surface and observe what does happen. We did this by allowing a periodic boundary condition (a feature of the general VIDSIM program[31] and of most computer programs for crystalline lattices) for Au_{Si} to translate those Au_{Si} that form just at the source side surface to the far side surface rather than keeping them on the proper surface or annihilating them. We continued to annihilate all Si_i that passed through either surface in the diffusion direction, x. This provided a source of Au_{Si} without a source of Au_i on the far side surface. As Si_i's arrived at the far surface they have a chance to displace the Au_{Si} and produce thereby Au_i. These Au_i could then either annihilate or diffuse into the bulk from the far side. With a little adjustment of Au_i injection rate at the source side and waiting for a proper concentration of Au_{Si} to develop, we were able to cause assumption (5) to become fulfilled. We found that the $[Au_{Si}](x)$ profile thus produced near the far side surface is indeed in accord with the theory of Gösele *et al.*[6] See Fig. 4. However, the profile near the source side surface is, as in proper simulations, like that in Schmid's thesis[33] and elsewhere, rather than as Gösele *et al.* predict. We maintain that this affirms two facts: a) the problem with the theory developed by Gösele and co-workers is in the initial assumptions and not in the deduction from those assumptions; and b) their is no fatal flaw in our computer programs.

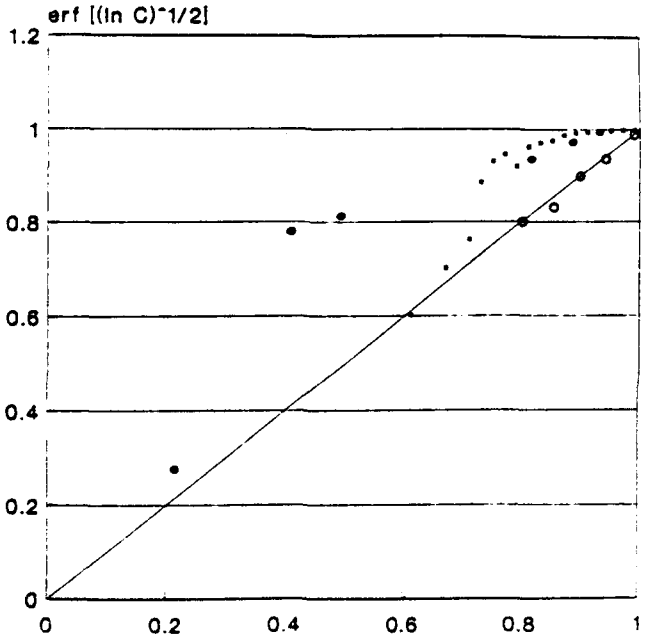


Fig. 4 — Result of direct simulation of kick-out mechanism modified from Fig. 3 in the following ways. a) The source is on one side only. b) The reflective boundary condition of the generic VIDSIM simulator, Ref. 31, has been retained so that Au_{Si}'s formed at the source side surface are artificially translated to the far side surface. This produces a source for Au_{Si} on the far surface with no corresponding source for Au_i. c) The injection rate has been adjusted and the time chosen so that the condition of "local equilibrium," Eq. (5) obtains near the far surface. Far surface data indicated by ○. The source side result is similar to Fig. 3.

Thus we conclude that the observation of a symmetric U shaped profile for Au_{Si} resulting from a Au source on one surface only compels the conclusion that the Au_i diffused all the way through the sample with negligible probability to displace a host Si atom even once. As noted at the beginning of this section, this requires that the energy of formation of Si_i must be too great for Si_i to play any role in thermal processes in Si.

We now pass on to consider the explanation of various surface treatment effects, which many have attributed to Si_i, without them.

II. OXIDATION INDUCED STACKING FAULTS AND DISLOCATION LOOPS BY VACANCY CLIMB

It is well known that oxide precipitates in Czochralski Si are generally surrounded by extrinsic stacking faults and dislocation loops.[34] There are often multiple planes of Si added to form these extrinsic faults. Careful study by electron microscopy reveal that they are often smooth and regular. It is also well established that there is local swelling of the Si when O atoms, which were disssolved as interstitials during crystal growth, converge form these oxide precipitates. The volume of the oxide exceeds that of the Si crystal which is consumed in its formation. The volume of the swelling concomitant with the dissolving of the O as interstitials is less than this difference between the Si crystal and the resulting oxide and is spread through the sample rather than localized.

The tendency of the SiO_2 precipitate to swell in the crystal lattice as it grows produces both tensile and compressive stress in the Si. The stress is compressive in the direction radial from the center of the precipitate and tensile in the direction tangential to the precipitate. The growth of extrinsic stacking faults with the concomitant dislocation loops relieves the tensile stress by producing extra planes of Si around the precipitate. This might occur either by precipitating Si_i's (if there were any Si_i's), which would have to be produced elsewhere, or by emitting V's from the core of the dislocation that surrounds the plane of the stacking fault.[35]

Thus, one can easily and naturally explain the formation of extrinsic stacking faults in Si by supposing the creation of vacancies along the core of the dislocations produced by the swelling of the precipitate. This is the locus of maximal stress. This leads one to expect the stacking faults to be more or less regular, symmetric and smooth, as is generally observed. If they were formed by precipitation of Si_i's, then one might expect them instead to be irregular and dendritic. A similar formation of extrinsic stacking faults is observed in metals where there is no question that V's completely dominate over self interstitials.[35,36]

One might confirm that oxide precipitation injects V's into the surrounding Si by positron annihilation techniques. Initial reports appear to do this.[37]

III. Si ION IMPLANT EFFECT ON DEEP B IMPLANTATION

Schreutlekamp *et al.* have reported[38] that the tendency of ion implanted B to straggle into Si, so that the doping profile is less abrupt than is desired, can be suppressed by ion implanting Si into the sample also. We recall that experiments that use electron irradiation to produce V + Si_i, Frenkel pairs, and electron spin resonance techniques to monitor the defect complexes that result have shown that Si_i's disappear from the sample too quickly to be observed (even at T = 2 K).[8] They also show that Si_i can displace substitutional B, B_{Si}, to drive the reaction

$$Si_i + B_{Si} \Rightarrow B_i, \tag{9}$$

while V form complexes with other vacancies (*i.e.* form divacancies and larger clusters) and with many types of impurity. Finally, the B_i is known to diffuse very rapidly.

The natural explanation of this is that the source of the problem is the tendency of B ion implantation to produce B_i's which diffuse to produce a straggle profile precisely because they cannot reverse Reaction (9). This because the enthalpy of formation of Si_i is larger than that of B_i.[4] In the absence of Si ion implantation, these B_i must diffuse too far to find a V which they can convert to the desired acceptor dopant via,

$$B_i + V \Rightarrow B_{Si}. \tag{10}$$

The Si ion implantation cure works because it produces the V's required for (10) so that the B_i can be stopped. Note that if there were any significant numbers of Si_i remaining from the Si ion implantation, they would enhance the B_i diffusion and make the B_{Si} profile less, rather than more, abrupt.

IV. OXIDATION AND NITRIDIZATION ENHANCED OR RETARDED DIFFUSION OF DONORS OR ACCEPTORS

It is well established that oxidation of a Si surfaces enhances the diffusion B and P beneath them while nitridization of these surfaces suppresses the diffusion of B and P.[39-42] On the other hand, Sb diffusion is affected to a lesser degree and in the opposite sense; *i.e.* Sb diffusion is suppressed by oxidation and enhanced by nitridization.

It is clear that oxidation of the surface produces a tensile stress in the Si tangential to surface, just as the formation of an oxide precipitate produces tensile stress in the bulk. Plainly, we should expect this tensile stress to be partly relieved by the injection of more vacancies from the incoherent interface between crystal and oxide into the Si crystal. This will enhance $[V]$ in the Si and thereby enhance the diffusion of species that diffuse primarily via exchange with V_{Si}. It is well established that V_{Si} pairs readily with both B and P.[8] (Recall that P_{Si}-V_{Si} pairs, "E-centers" are observed in P doped Si that has been rapidly quenched[43] from temperatures near, but below, the melting point.) This is partly because of a coulomb attraction between the amphoteric V_{Si}, which assumes a V_{Si}^{2+} ionization state in p-type material (*i.e.* where the B is) but assumes a V_{Si}^{-} or V_{Si}^{2-} ionization in n-type material (*i.e.* where the P is). Thus, the oxidation enhancement of B and P diffusion is easily understood in terms of the well established properties and interactions of V_{Si}.

It is also clear that nitridization of the Si surface stresses the crystal in the opposite sense, *i.e.* puts it into compression. It is natural to suppose that this compression inhibits the generation of V's which implies that $[V]$ is suppressed. Thus, one should expect the vacancy mediated diffusion of P and B to be suppressed, as is observed.

So far the discussion of this section has proceeded as if the diffusivity of impurities were simply proportional to $[V]$. It should be evident that is almost certainly an oversimplification. The coulomb interaction between ionized impurities and V_{Si} with its five ionization states within the bandgap and the observation of recombination enhanced diffusion for V_{Si}, B and several complexes imply a more complicated relation.[30]

Reports that Sb diffusion is suppressed by oxidation and enhanced by nitridization,[41,42] combined with the arguments that oxidation must increase $[V]$ and that Si_i cannot play a significant role, reenforce the supposition that we really must consider the relation between $[V]$ and impurity, X, diffusivity, $D(X)$, more carefully.

We first note that the report of a depression of D(Sb) with oxidation is based solely on determination of the depth of the p-n junction formed when Sb is diffused into p-type Si. We know of no measurement of the effect of oxidation or nitridization upon D(Sb) in homogeneous material. Recall that we already know that, on account of the coulomb interaction between ionized impurities and ionized vacancies, the vacancy mediated diffusivity of donors is enhanced in n-type material and suppressed p-type material. This occurs because of the attraction between D^+ and V^{2-} or V^- in n-type material and because of the repulsion between D^+ and V^{2+} in p-type material. This coulomb enhancement changes to coulomb repulsion at a point very close to the nominal p-n junction. Now, because $[V]$ is well above 10^{16} cm^{-3} at the temperatures in question even without oxidation enhancement and because the amphoteric V's always tend to compensate the doping of the sample, one additional consequence of increasing $[V]$ by surface oxidation is to enhance this compensation. This causes the point at which coulomb enhancement converts to suppression of D(Sb) to move toward the surface. Assuming that the coulomb interaction between Sb^+ and V_{Si} plays a major

role in determining D(Sb), we propose it is this displacement of the compensation point during diffusion toward the surface by oxidation that accounts for the reduction of the depth of the resultant junction. We predict that D(Sb) within the n-type region is in fact enhanced, as could be verified by radiotracer experiments.

Finally, one may wish to consider observations of the growth and shrinkage of stacking faults on one side of a sample due to the oxidation of the other side. Such observations have been reported[44] by Taniguchi *et al.* who used Si_3N_4 and thick polycrystalline Si films to prevent oxidation and buffer the one side of the sample from the stress of the Si_3N_4. It is found that for samples of normal thickness (*e.g.*, 0.5 mm) stacking faults previously grown on the masked side will shrink when the far side is oxidized. They report that the rate of this shrinkage has an activation energy of 4.9 eV, which is roughly consistent with that for Si self diffusion. However, for thin samples (10 to 80 μm) a few of the stacking faults will start to grow longer (not deeper) after some thickness dependent incubation time; the other stacking faults shrink. They also report that the stacking faults grow deeper at a rate $dZ/dt = 5 \exp(-2.6\ \text{eV/kT})\ m/s$ when oxidation occurs.

Taniguchi *et al.* interpreted[44] their observations under the assumptions that Si_i processes completely dominated, that the shrinkage of stacking faults occurs by emission of Si_i's into the bulk, and that elongation of the stacking faults occurs by absorption of Si_i's that have been injected at above thermal equilibrium concentrations by the oxidation of the far surface. They adjusted five parameters to fit their observations of the length of the longest stacking faults under these assumptions. They concluded that the activation energy for Si_i migration is 4.0 eV and that the enthalpy of formation of Si_i is only 0.7 eV while its entropy of formation is -7.6 k, *i.e.* negative implying decrease of disorder upon forming the defect. These conclusions are in gross contradiction of the parameterizations used to fit other properties by those who assert the importance of Si_i. (See, *e.g.* Ref. 12.) They imply that Si_i should be easy to retain from a variety of processes in concentrations that would make it easy to observe, which has never occurred. They imply that the equilibrium concentration of Si_i at 1300 K is 4.6×10^{16} cm^{-3}. Recall that Okada finds[20] the concentration of vacancies to be 2×10^{16} cm^{-3} greater than any concentration of Si_i.

We regard these contractions as further indictment of the assumption that Si_i plays any major role in thermal processes. (We hold that the clearest indictment is the inability to explain a two sided profile of Au_{Si} resulting from a source on one side only.) We suggest that a better interpretation of observations such as those reported by Taniguchi *et al.* begins with the observation that the incubation period before some stacking faults begin to elongate corresponds, according to their formula, to the time required for the depth of the stacking fault to reach the far surface. When this happens, a large part of the dislocation that bounds the stacking fault disappears into the interface on the far side. This produces a large reduction in the total energy of the stacking fault and stabilizes it. However, this occurs only for the thin samples; oxidation of the far side, with its concomitant tensile stress, injects V's as described above, which annihilate extrinsic stacking faults by causing their dislocations to climb toward the near surface. Thus, the excess energy of the stacking faults in thick samples, where the faults do not reach the far side, is relieved by (vacancy mediated) self diffusion. In the thin (10 to 80 μm) samples, once some of the stacking faults have extended from one side of the sample to the other, they can further reduce the energy of the sample by consuming those stacking faults that do not extend across the sample. The sample energy can be reduced even further by combining stacking faults that extend across the sample. Both annealing processes can be accomplished by glide of the dislocations that thread from one interface to the other and which bound the stacking fault laterally, as well as by climb mediated by vacancy emission and absorption at the cores of the dislocations. As these stacking faults cannot get any deeper than the width of the sample, they must elongated in this annealing process. Given that the density of stacking faults in the thin samples is of order 2×10^5 cm^{-2}, the spacing between their centers is about 45 μm. Their length is initially 10 μm and grows to more than 60 μm. We believe that annealing by dislocation glide is a natural explanation for this elongation.

REFERENCES

1. A. Seeger and K. P. Chik, Phys. Status Solidi *29*, 455 (1968).
2. B. Tuck, Introduction to Diffusion in Semiconductors, IEE Monograph 16 (Peter Peregrinus Ltd., Stevenage, England, 1974).
3. J. C. Bourgoin and J. W. Corbett, Radiat. Eff. *36*, 157 (1978).
4. J. A. Van Vechten, Handbook on Semiconductors, Vol. 3, ed. S. P. Keller (North Holland, Amsterdam, 1980) p. 1.
5. A. Seeger, Phys. Status Solidi A *61*, 521 (1980).
6. U. Gösele, W. Frank and A. Seeger, Appl. Phys. *23*, 361 (1980).
7. D. Mathiot and J. C. Pfister, J. Appl. Phys. *55*, 3518 (1984).
8. G. D. Watkins, Deep Centers in Semiconductors, ed. S. T. Pantelides (Gordon Beach, New York, 1985) p. 147.
9. J. A. Van Vechten, J. Electron. Mater. *14a*, 293 (1985).
10. K. C. Pandey, Phys. Rev. Lett. *57*, 2287 (1986).
11. M. Yoshida, Jpn. J. Appl. Phys. *27*, 967 (1988).
12. W. Taylor, B. P. R. Marioton, T. Y. Tan and U. Gösele, Radiat. Effects a. Defects in Solids *111*, 131 (1989).
13. J. A. Van Vechten, J. Schmid and N. C. Myers, Defect Control in Semiconductors (Proc. Int. Conf. Sci. Tech. Defect Control in Semicond., Yokohama 1989). ed. K. Sumino (North Holland, Amsterdam, 1990) p. 41.
14. U. Schmid, J. A. Van Vechten, N. C. Myers and U. Koch, Mat. Res. Soc. Symp. Proc. *163*, 609 (1990).
15. S. Dannefaer, S. Kupca, B. G. Hogg and D. P. Kerr, Phys. Rev. *B 22*, 6135 (1980); see also S. Dannefaer, N. Fruensgaard, S. Kupca, B. G. Hogg and D. P. Kerr, Can. J. Phys. *61*, 451 (1983).
16. W. Brown and M. Augustyniak, J. Appl. Phys. *30*, 1300 (1959); H. Flicker, J. J. Loferski and J. Scott-Monck, Phys. Rev. *128*, 2557 (1962).
17. W. Kohn, Phys. Rev. *94A*, 1409 (1954).
18. F. Seitz and J. S. Koehler, Solid State Physics, vol. 2, eds.

F. Seitz and J. S. Koehler (Academic Press, New York, 1956), p. 305.
19. S. Dannefaer, P. Mascher and D. Kerr, Phys. Rev. Lett. *56*, 2195 (1986).
20. Y. Okada, Phys. Rev. *B 41*, 10741 (1990).
21. See, *e.g.* R. O. Simmons and R. W. Ballufi, Phys. Rev. *125*, 862 (1962).
22. J. A. Van Vechten, Phys. Rev. *B 33*, 2674 (1986).
23. F. Demond, S. Kalbitzer, H. Mannsprenger and H. Damjantschitsch, Phys. Lett. *A 93*, 503 (1983).
24. W. R. Wilcox, T. J. LaChapelle and D. H. Forbes, J. Electrochem. Soc. *111*, 1377 (1964).
25. F. A. Huntley and A. F. W. Willoughby, Solid-State Electron. *13* 1231 (1970).
26. F. A. Huntley and A. F. W. Willoughby, J. Electrochem. Soc. *120*, 414 (1973).
27. S. Coffa, L. Calcago, S. U. Campisano, G. Calleri and G. Ferla, J. Appl. Phys. *64*, 6291 (1988).
28. F. C. Frank and D. Turnbull, Phys. Rev. *104*, 617 (1956).
29. M. D. Sturge, Proc. Phys. Soc. (London) *73*, 297 (1957).
30. J. A. Van Vechten, Phys. Rev. *B 38*, 9913 (1988) and therein.
31. U. Schmid, N. C. Myers and J. A. Van Vechten, Comp. Phys. Commun. *58*, 329 (1990).
32. E. R. Weber: Properties of Silicon, EMIS Datareview (INSPEC, IIE, London, 1988).
33. U. Schmid, Thesis, Dept. of Physics, Oregon State University (1988).
34. W. Zulander and D. Huber, Czochralski-Grown Silicon (Springer, Berlin, 1982), *e.g.* p. 71.
35. J. W. Matthews and J. A. Van Vechten, J. Cryst. Growth *35*, 343 (1976).
36. W. C. Dash, J. Appl. Phys. *31*, 2275 (1960).
37. W. Puff and S. Dannefaer, ICSDS-4 (London, 7 - 8/1990) to be published.
38. R. J. Schreutelkamp, W. X. Lu, F. W. Saris, K. T. F. Janssen, J. J. M. Ottenheim, R. E. Kaim and J. F. M. Westdorp, Mater. Res. Soc. Symp. Proc. *157* 691 (1990).
39. A recent review of this literature is: P. M. Fahey, P. B. Griffin, and J. D. Plummer, Rev. Mod. Phys. *61*, 289 (1989).
40. S. Mizuo and H. Higuchi, J. Electrochem. Soc. *129*, 2292 (1982).
41. S. Mizuo and H. Higuchi, J. Electrochem. Soc. *130*, 1942 (1983).
42. S. T. Ahn, H. W. Kennel, J. D. Plummer and W. A Tiller, Appl. Phys. Lett. *53*, 1593 (1988).
43. A. Chantre, M. Kechouane and D. Bois, Physica *116B*, 547 (1983).
44. K. Taniguchi, D. A. Antoniadis and Y. Matsushita, Appl. Phys. Lett. *42*, 961 (1983).

PCSI-19 1/92

CHARACTERIZATION OF POINT DEFECT GENERATION AT SILICON SURFACES USING GOLD DIFFUSION

R.K. Graupner[1,2], J.A. Van Vechten[1] and P. Harwood[3]

1) Oregon State University, Corvallis, OR 97331-3211
2) Now with Komatsu Silicon U.S.A., Santa Clara, CA 95051
3) Wacker Siltronic, Portland, OR 97283-0180

Abstract

We demonstrate the utility of observations of Au diffusion for the characterization of point defect generation and annihilation rates at surfaces and of vacancy distributions and diffusivity in the substrate. We note particularly the effects of ambient gases and surface finish on the diffusion process and on the resultant distributions of electrically active Au. We deposited Au on commercial float zone Si in a vacuum system after the Si had reached the diffusion temperature (1233 K) and had been annealed in various ways. Contrary to a previously published report, we find the electrically active Au with a one-sided profile when the Au is deposited and annealed in a vacuum. We obtain the previously reported two-sided profiles for electrically active Au when millitorr levels of O_2 are present in the ambient gas during diffusion or during a preanneal prior to Au deposition. We conclude that the polished or etched silicon surfaces lack the imperfections needed to make them effective sources or sinks for vacancies or self-interstitials, but that surface roughening caused by O_2 induced evaporation of SiO is sufficient to create very effective sources or sinks.

I. INTRODUCTION

The diffusion of gold into silicon has been the subject of extensive experiment, theory and controversy [1,2,3,4,5,6]. Previously reported experiments show a distinctive U-shaped profile for electrically active, substitutional gold, Au_{Si}, which results even when the source is on one side only and precautions are taken to avoid surface diffusion. It is generally agreed that the Au diffuses rapidly as an interstitial, Au_i, and that the concentration of Au_i, is orders of magnitude less than the equilibrium concentration of Au_{Si}. There is controversy concerning how Au_i becomes Au_{Si} with a symmetric profile from a one-sided source. Initially a Frank-Turnbull (FT) or dissociative mechanism (1) was invoked to explain this profile. In this mechanism vacancies, V, generated at the surfaces diffuse into the bulk and combine with Au_i to form Au_{Si}.

$$Au_i + V \rightleftharpoons Au_{Si} \qquad (1)$$

Others claim that the profile results from a "kick-out" mechanism and is determined by the creation of silicon self-interstitials, Si_i, in the bulk via reaction (2) and their subsequent out-diffusion and annihilation at the wafer surfaces.

$$Au_i \rightleftharpoons Au_{Si} + Si_i \qquad (2)$$

In both of these models it was assumed that the sample surface is an ideal source or sink for vacancies and interstitials, that

the interior of the sample is at thermodynamic equilibrium at the onset of the diffusion (time t=0) and that point defect creation or annihilation at the surfaces is the key factor which creates a two-sided profile.

Van Vechten, Schmid, and Zhang have shown through direct Monte Carlo simulation [5] that the kick-out hypothesis cannot produce the reported two-sided and nearly symmetric profile from a strictly one-sided source starting with a perfect sample. Simulations of the FT mechanism for the corresponding assumptions, to be published elsewhere, are two-sided but not symmetric and do not resemble the reported profiles well either.

The initial motivation for this study of Au diffusion into Si was to gain an understanding of how the previously reported U-shaped profile can result from a one-sided diffusion and to make use of the rapidly diffusing Au interstitials to plot out the point defect profiles and to determine their diffusion rates. We noted that because the Au interstitials diffuse so fast, we can introduce them at various times after the sample temperature, T, has been raised to levels that self diffusion is significant and use them to decorate vacancies rendering these immobile in a short time. In this way we sought to make "snapshots" of the vacancy distribution during the high T processes by observing the distribution of Au_{Si} by spreading resistance profiling, INAA, and TEM. We also noted that the kick-out model and the FT model make qualitatively different predictions for the case of a sample preannealed prior to the deposition of the Au. According to the FT model, the profile

should be almost entirely determined by the time the vacancies have to diffuse with or without the Au because the profile is determined by the distribution of sites where the Au_i's find vacancies to occupy. According to the kick-out model, the profile is determined by the time the Au is present at high T because the profile results from the interplay between the Au_{Si} and the Si_i that they produce.

We report here our conclusions on the effect of ambient gas on surface perfection and the resulting effectiveness of the surface as a source or sink for point defects.

II. EXPERIMENT

Commercially available float-zoned 100mm diameter boron-doped silicon wafers of <100> orientation, resistivity of 60-70 ohm-cm were cleaved into strips 1 cm x 9 cm x .05 cm prior to processing. This material has a low interstitial oxygen concentration and is free of dislocations. Samples 1 - 12 were mirror polished on one side and caustic etched on the other side. The caustic etching is sufficient to remove all surface damage as revealed by dislocation etches. Samples 13-17 were mirror polished on both sides. No surface preparation or cleaning was performed. Recent work indicates that the presence of a native oxide layer is essential to the diffusion results reported here.

Figure 1 shows the apparatus used for the sample annealing and the gold deposition. An Ohmic contact was made to both the silicon sample and heater strip at both ends with graphite electrodes. An ac voltage controlled by an autotransformer was applied to both the sample and heater strip to provide resistance heating. The sample was radiatively heated by the heater strip, a 0.02 ohm-cm boron doped silicon strip mounted 1 mm from the sample, until the sample conductivity was sufficient to permit resistance heating. The side of the heater strip facing the sample was coated with an SiO_2 layer to prevent contamination of the sample with out-diffused boron. An optical pyrometer mounted outside the vacuum chamber measures the temperature of the heavily doped strip and controls the autotransformer. The optical pyrometer is calibrated at the melting

point of silver (1233 K). The sample is mounted in a vacuum chamber above a resistance heated molybdenum boat containing the gold to be evaporated onto the sample. This equipment allows gold deposition onto the heated sample at any time during the annealing while maintaining the sample at processing temperature. During annealing, a gold layer with a calculated thickness of 3 nm was deposited on samples 1-11, and a 60 nm layer was deposited on samples 12-17. The ambient gas was established by venting the chamber with argon, with air and by operating in a vacuum of <4.0 E-6 Torr.

After processing, the samples were bevel polished and measured by a spreading resistance probe (SRP). The spreading resistance values were converted into resistivity by use of three p-type calibration samples covering the range 22-118 ohm-cm. The resistivity change after gold diffusion is used to calculate the concentration of electrically active substitutional gold using the method discussed by Coffa et al [3]. We used a degeneracy factor of 28 [7] to facilitate comparison of the Au profiles presented here with the profiles published by Coffa et al [3]. INAA applied to sample 13 gave total Au concentrations which agreed well with the total electrically active gold concentration measured with the SRP. The large degeneracy factor may mask the presence of non-electrically active gold in the sample since the degeneracy factor calculations [7,2] assume that virtually all of the Au diffused into the sample occupies electrically active Au_{Si} sites. Non-electrically active gold could result from the precipitation of Au_i into vacancy clusters in the interior of the sample.

EFFECT OF AMBIENT GAS ON GOLD DIFFUSION PROFILE

Figure 2 shows the measured profile for the sample 13 which was annealed in an actively pumped vacuum (<4.0E-6 Torr) for 30 minutes at 1233 K following the deposition of the gold. In all figures the gold was deposited on the surface defined by depth=0.In stark contrast to the two-sided profiles published by Coffa [3], or Stolwijk et al [2], this profile is clearly one-sided. All samples which were annealed in vacuum, including those with caustic etched backsides, show a similar profile.

Figure 3 shows the measured profile for sample 9 which was annealed at atmospheric pressure in argon at 1233 K. This sample has a two-sided or U-shaped profile but is asymmetric. The lower Au_{Si} concentrations compared to sample 13 are probably due to the smaller amount of gold which was deposited on the surface.

Figure 4 compares the effects of 5 min. Au diffusions at 1233 K in air at 0.5 Torr pressure with (sample 15) and without (sample 16) a 30 min. preanneal in the same ambient gas at 1233 K. After preannealing sample 15, the chamber was evacuated, Au deposited, pressure returned to 0.5 Torr and the sample annealed for an additional 5 min. without any change in T. The Au_{Si} profile is U-shaped but asymmetric and similar in appearance to the sample 9 shown in figure 3. This similarity occurs even though the Au was diffused into sample 15 for only 5 min. as compared to 30 min. diffusions for samples 9 and 13. Sample 16 which received the 5 min. Au diffusion without a preanneal, shows a much lower concentration of Au_{Si}.

DISCUSSION

For either the FT or kick-out mechanisms, the two-sided profile is caused by the wafer surfaces acting as sources or sinks for the point defects which mediate the diffusion of the Au_{Si}. The one-sided profiles obtained when the diffusion occurs in a vacuum indicate that undamaged polished or etched silicon surfaces consist primarily of steps, lacking the energetically costly kink sites and are ineffective as sources or sinks for point defects. Swartzentruber et al [8] used STM to determine step and kink energies on (100) silicon surfaces and noted that their analysis was "complicated by the fact that there are so few kinks.". The deposition of gold and the subsequent formation of a gold eutectic layer creates the required kink sites only on the side receiving the Au deposition, resulting in the observed one-sided Au_{Si} profiles.

For the samples annealed in argon or reduced pressure air a different mechanism is responsible for the formation of the kink sites. Annealing silicon wafers in inert gases is known to cause "thermal etching" of the surfaces due to O_2 partial pressures below the 10^{-6} atm. range in the annealing gas [9,10,11]. The O_2 attacks the silicon surface to form SiO which evaporates from the surface. The resulting surface roughening and etch pit formation provide ample kink sites to drive the point defect reactions (1) or (2). At higher oxygen partial pressures, a protective layer of SiO_2 forms which protects the surface from damage. The occurrence of the two-

sided profile in both argon at atmospheric pressure and air at reduced pressure demonstrates that thermal nitridation of the surface is not the primary reaction which creates the kink sites. Preliminary evidence suggests that surfaces oxidized at 1233 K in air at atmospheric pressure also lack the sites necessary to sink or source point defects. We suggest that annealing samples in inert gases with low partial pressures of O_2 is an effective method for injecting vacancies into the sample.

The anneals described in the literature occurred in nitrogen [3], argon [1,2] or vacuum [1]. However, the vacuum anneals in [1] occurred in evacuated quartz ampoules as opposed to actively pumped vacuum systems. In all of these cases it is reasonable to assume that oxygen contamination occurred at ppm level and resulted in thermal etching which created the sinks or sources resulting in a two-sided profile. Gas phase transfer of Au from the frontside to the backside would be expected to occur in the ampoule annealed samples since the vapor pressure of Au at 1233 K is 1 E-6 Torr. Such transfer could account for the symmetric profiles reported by Stolwijk et al [1,2].

The results from samples 15 and 16 suggest that the presence of gold is not solely responsible for the development of two-sided point defect profiles but also serves to decorate the point defect concentration profiles present at the time it is diffused into the wafer. The difference between the profiles of samples 15 and 16 is predicted by the FT mechanism but is inconsistent with the solely interstitial based kick-out model used by Coffa et al [3].

According to such a model the Au_{Si} profiles of samples 15 and 16 would be predicted to be identical since there is no out-diffusion of Si_i prior to the creation of high concentrations of these point defects by the deposition and subsequent diffusion of the gold. Given this assessment of the role of gold as a defect marker, the techniques described in this paper will serve as a useful characterization tool for studying the generation of point defects at surfaces and in the bulk.

ACKNOWLEDGEMENTS

This work was supported in part by U.S. AFOSR-89-0309.

The authors wish to acknowledge T. Monson for his assistance in the sample processing, L. Ungier for his suggestions on the design of the vacuum annealing system and T. Nebert for the SRP measurements.

REFERENCES

1. N.A. Stolwijk, B. Schuster and J. Holzl, Appl. Phys., **A33**, 133 (1984)
2. N.A. Stolwijk, J. Holzl and W. Frank, Appl. Phys., **A39**, 37 (1986)
3. S. Coffa, L. Calcagno, S.U. Campisano, G. Calleri and G. Ferla, J. Appl. Phys., **64**, 6291 (1988)
4. U. Gosele, W. Frank and A. Seeger, Appl. Phys., **23**, 361 (1980)
5. J.A. Van Vechten, U. Schmid, and Zhang Q.-S., J. Electron. Materials, **20**, 431 (1991)
6. W.R. Wilcox and T.J. LaChapelle, J. Appl. Phys., **35**, 240 (1964)
7. S. Coffa, G. Calleri, L. Calcagno, S.U. Campisano and G. Ferla, Appl. Phys. Lett., **52**, 558 (1988)
8. B.S. Swartzentruber, Y.-W. Mo, R. Kariotis, M.G. Lagally and M.B. Webb, Phys. Rev. Lett., **65**, 1913 (1990)
9. C. Wagner, J. Appl. Phys., **29**, 1295 (1958)
10. R.V. Giridhar and K. Rose, J. Electrochem. Soc., **135**, 2803 (1988)
11. J.R. Monkowski, Microcontamination, **Feb/Mar**, 37 (1984)

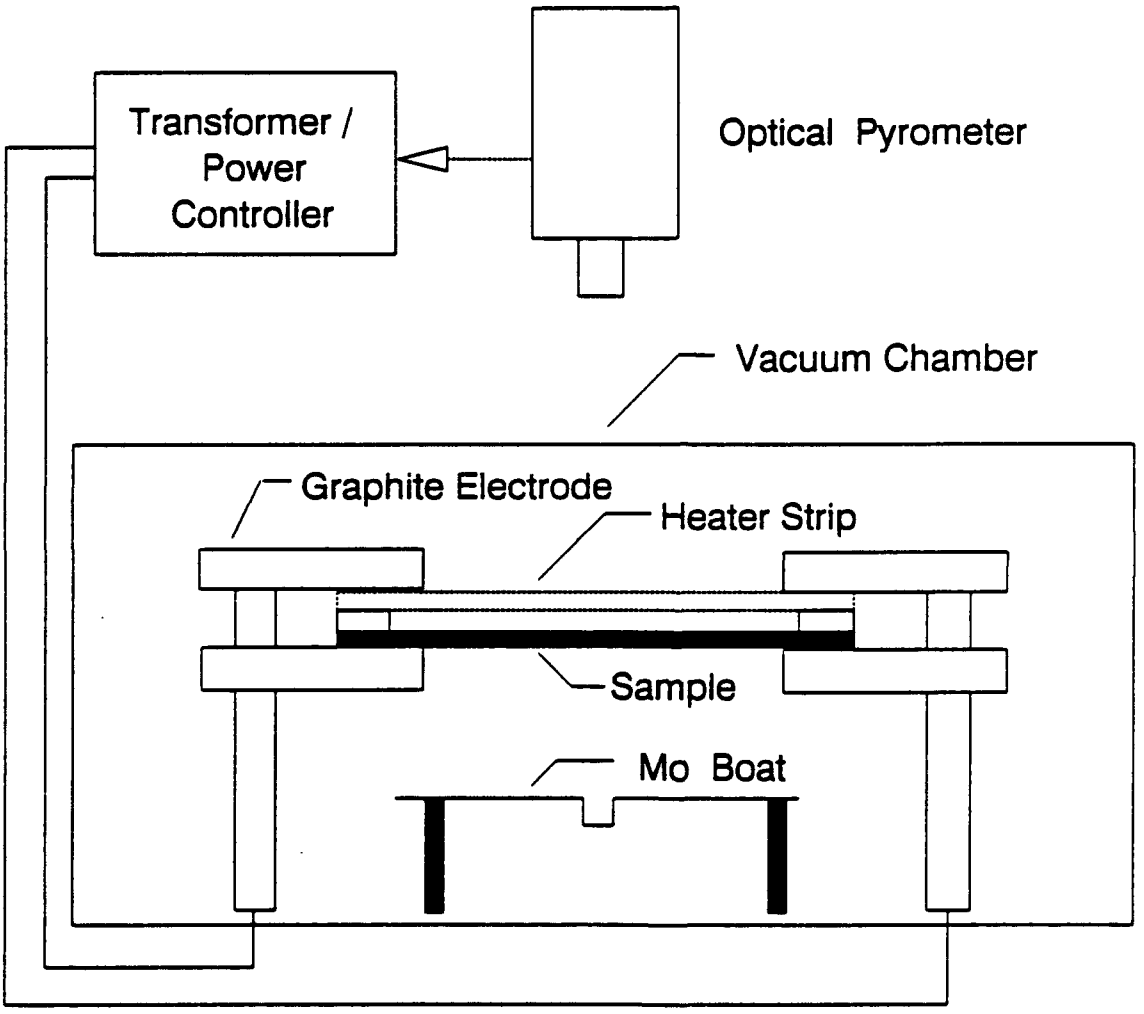
FIGURE CAPTIONS

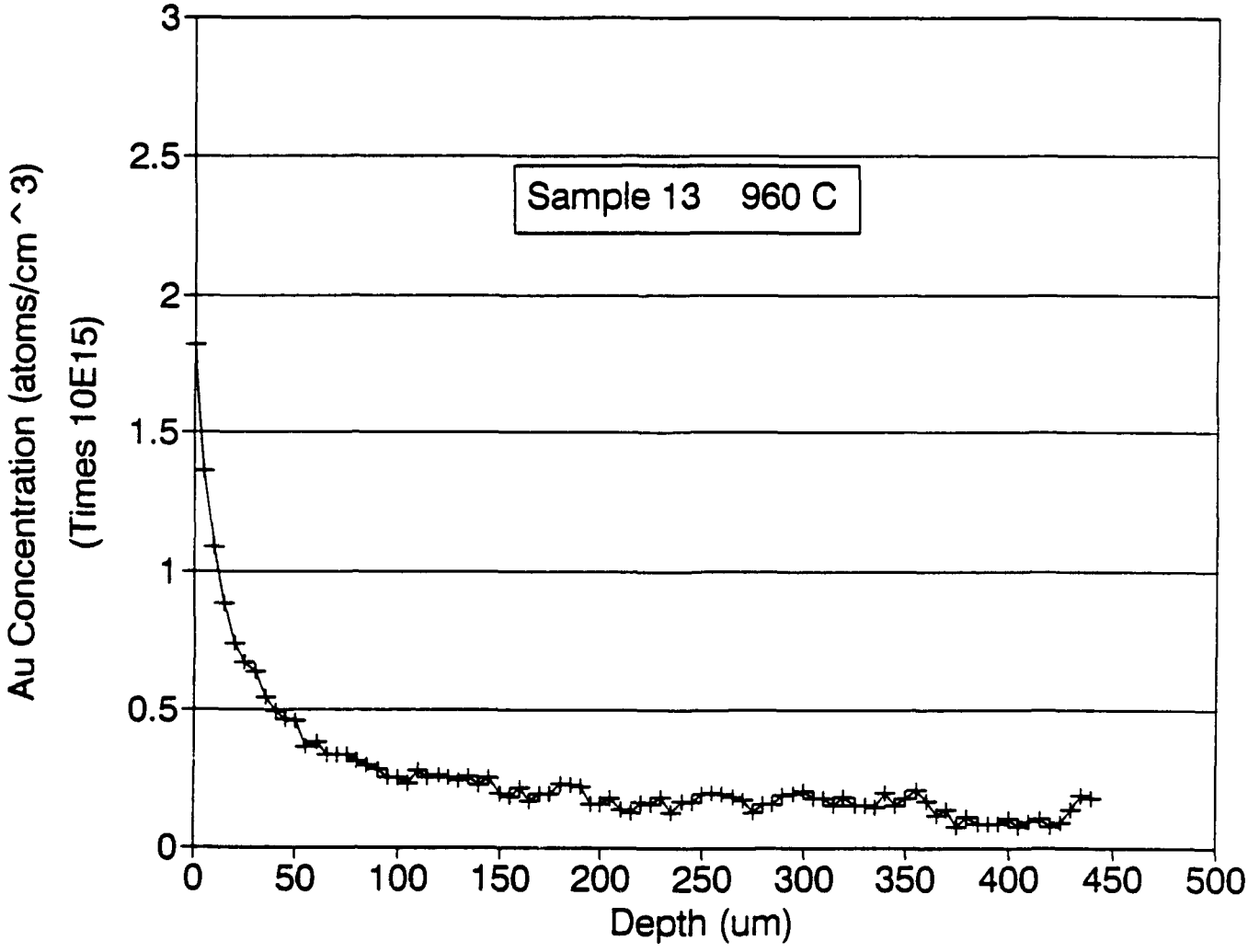
1. Schematic diagram of the apparatus used for gold deposition and annealing.

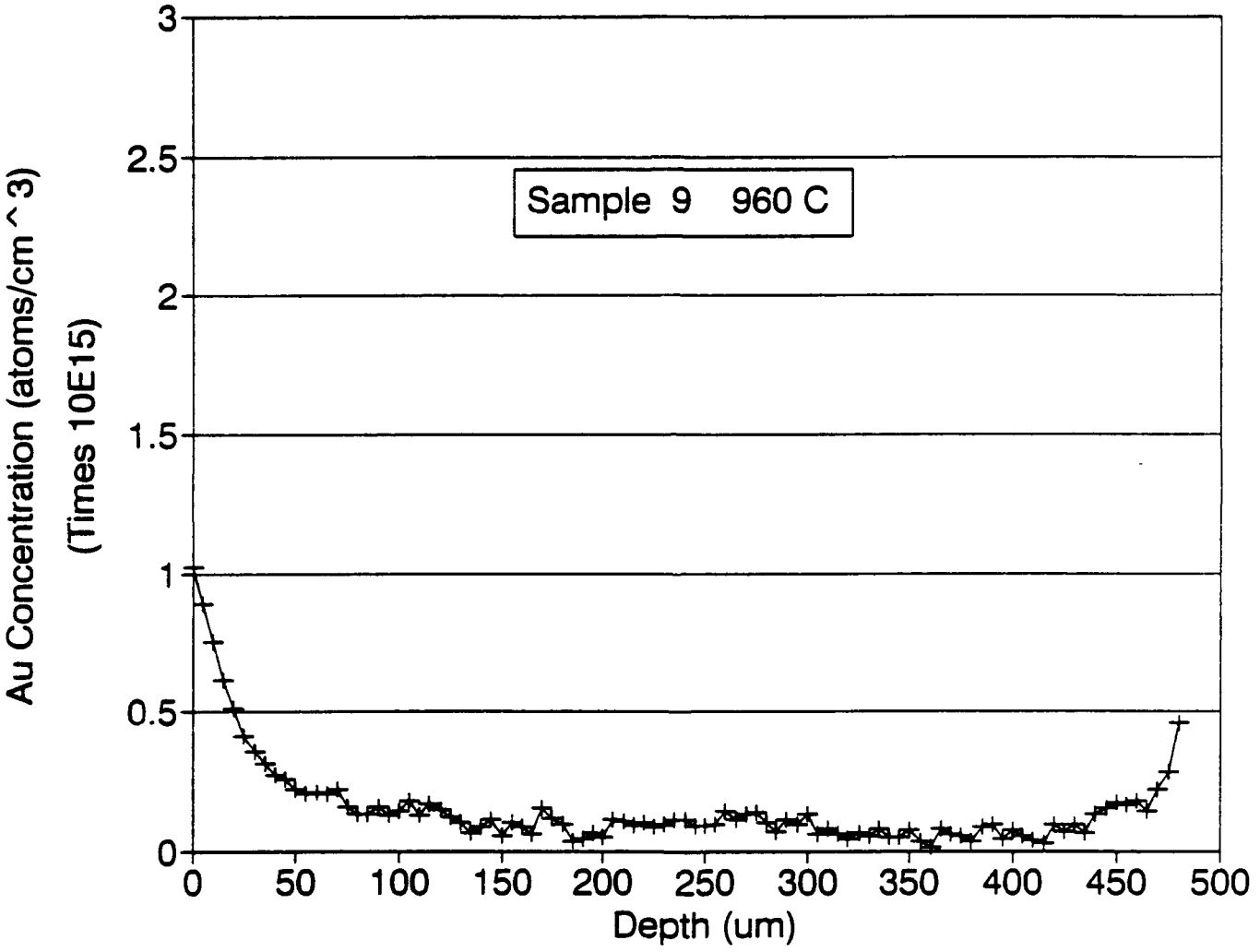
2. Electrically active Au_{Si} concentration for sample 13 with no preanneal, annealed for 30 min. at 1233 K in vacuum after gold deposition.

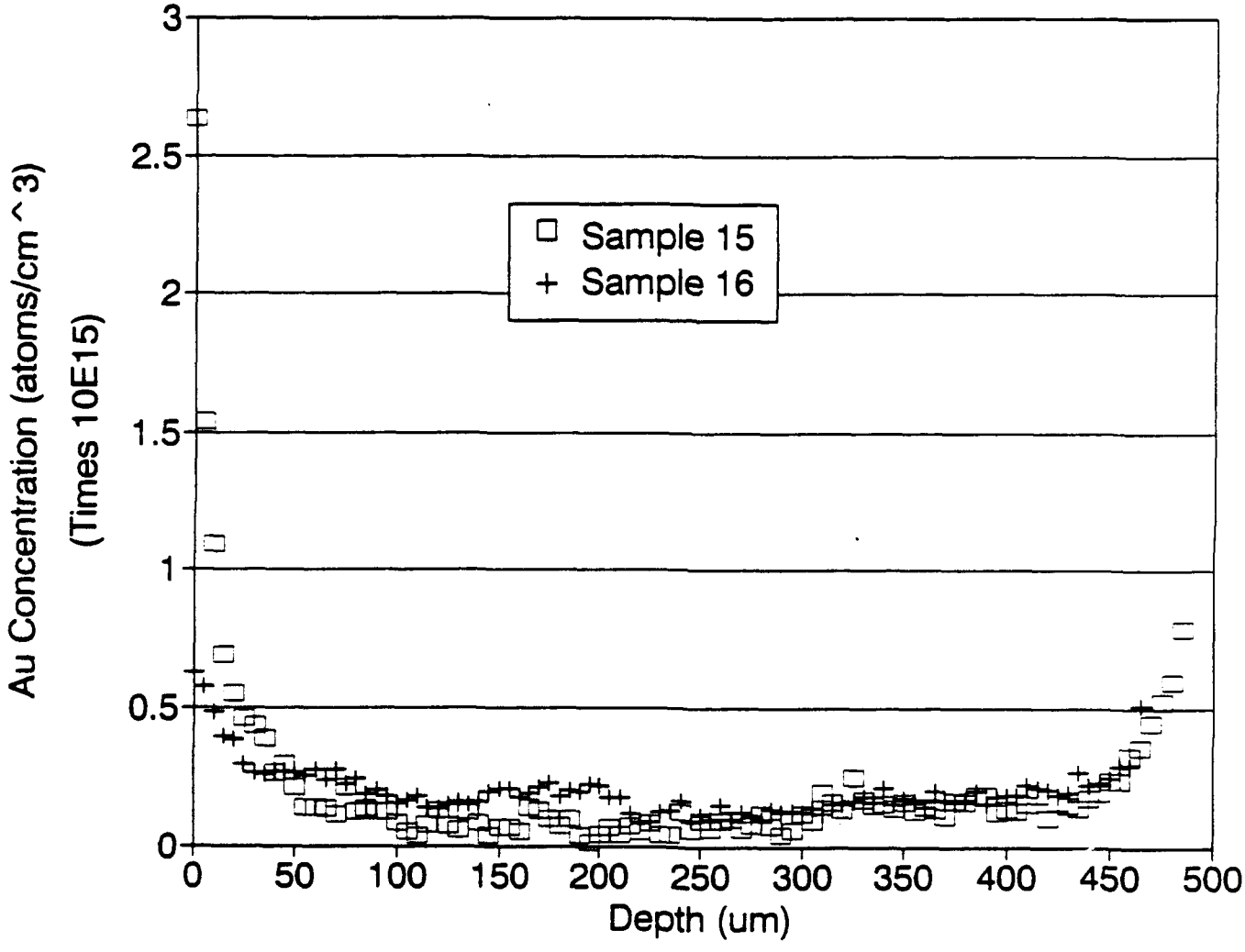
3. Electrically active Au_{Si} concentration for sample 9 with no preanneal, annealed for 30 min. at 1233 K in 760 Torr argon after gold deposition.

4. Electrically active Au_{Si} concentrations for sample 15 with a 30 min. preanneal at 1233 K in 500 mTorr air, annealed for 5 min. at 1233 K in 500 mTorr air after gold deposition and sample 16 annealed for 5 min. at 1233 K in 500 mTorr air after gold deposition without a preanneal.


Transformer / Power Controller
Optical Pyrometer
Vacuum Chamber
Graphite Electrode
Heater Strip
Sample
Mo Boat


Sample 13 960 C
Au Concentration (atoms/cm ^ 3)
(Times 10E15)
Depth (um)
0
0.5
1
1.5
2
2.5
3
0
50
100
150
200
250
300
350
400
450
500


Sample 9 960 C
Au Concentration (atoms/cm ^ 3)
(Times 10E15)
Depth (um)
3
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
300
350
400
450
500


Au Concentration (atoms/cm ^ 3)
(Times 10E15)
3
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
300
350
400
450
500
Depth (um)
Sample 15
Sample 16

THE EFFECT OF VACANCIES GROWN INTO SILICON ON GOLD DIFFUSION

R.K. Graupner *, J.A. Van Vechten**, P. Harwood***,
and T.K. Monson**
* Komatsu Silicon U.S.A., Santa Clara, CA 95051
** Oregon State University, Corvallis, OR 97331-3211
*** Wacker Siltronic, Portland, OR 97283-0180

ABSTRACT

We propose a generalized model for gold diffusion in silicon based on the effect of the high concentrations of vacancies and vacancy complexes in the as-grown silicon. The monovacancy profiles calculated using this model are identical to the substitutional gold profiles calculated using the kick-out model. We deposited Au on commercial float zone Si in a vacuum system after the Si had reached the diffusion temperature (1233 K) and had been annealed in various ways. Contrary to previously published reports, we find the electrically active Au with a nearly one-sided profile when the Au is deposited on samples which were preannealed in vacuum. We conclude that annealed silicon surfaces lack the imperfections needed to make them effective sources or sinks for vacancies or self-interstitials. We propose that this can cause a high degree of supersaturation in the as-grown silicon crystal since the point defects cannot annihilate at the surfaces to maintain equilibrium as the crystal is cooled.

INTRODUCTION

The diffusion of gold into silicon has been the subject of extensive experiment, theory and controversy [1,2,3,4,5,6]. Previously reported experiments show a distinctive U-shaped profile for electrically active, substitutional gold, Au_{Si}, which results even when the source is on one side only and precautions are taken to avoid surface diffusion. It is generally agreed that the Au diffuses rapidly as an interstitial, Au_i, and that the concentration of Au_i, is orders of magnitude less than the equilibrium concentration of Au_{Si}. There is controversy concerning how Au_i becomes Au_{Si} with a symmetric profile from a one-sided source. Initially a Frank-Turnbull (FT) or dissociative mechanism (1) was invoked to explain this profile. In this mechanism vacancies, V, generated at the surfaces diffuse into the bulk and combine with Au_i to form Au_{Si}.

$$Au_i + V \leftrightarrows Au_{Si} \qquad (1)$$

The FT mechanism with the assumption that the vacancies are only generated at the surfaces has not yet been shown to explain the profiles observed in dislocation-free silicon.

Van Vechten, Schmid, and Zhang have shown through direct Monte Carlo simulation that the FT mechanism cannot produce the reported nearly symmetric profile from a strictly one-sided source starting with a perfect sample. These simulations, to be published elsewhere, are two-sided but not symmetric and differ

significantly from the reported profiles.

Others claim that the profile results from a "kick-out" (KO) mechanism and is determined by the creation of silicon self-interstitials, Si_i, in the bulk via reaction (2) and their subsequent out-diffusion and annihilation at the wafer surfaces.

$$Au_i \rightleftharpoons Au_{Si} + Si_I \qquad (2)$$

Both of these models assume that the sample surface is an ideal source or sink for vacancies and interstitials, that the interior of the sample is at thermodynamic equilibrium at the onset of the diffusion (time t=0) and that point defect creation or annihilation at the surfaces is the key factor which creates a two-sided profile. The KO model also assumes that a uniform concentration of Au_i is established at the start of the diffusion process.

Van Vechten, Schmid, and Zhang have shown through direct Monte Carlo simulation [5] that the kick-out hypothesis cannot produce the reported two-sided and nearly symmetric profile from a strictly one-sided source starting with a perfect sample. These simulations demonstrate that the assumption of a uniform Au_i is incorrect. Nevertheless, the numerical solution of the partial differential equation resulting from the KO mechanism provides a remarkable fit to the experimental data. Accordingly, we sought models for Au diffusion based on vacancy and vacancy complexes which would produce the same differential equation as the KO model, but did not require the existence of significant concentrations of silicon self-interstitials.

A GENERALIZED MODEL FOR GOLD DIFFUSION

We propose that a generalized model for gold diffusion be considered. The reaction

$$A \rightleftharpoons B + C \qquad (3)$$

is assumed to establish a local equilibrium governed by the equation:

$$\frac{[B][C]}{[A]} = K \qquad (4)$$

A = A homogeneously distributed defect in the silicon at t=0 which is not decorated by gold. To a first approximation, [A] is assumed constant in time.

B= A defect which is capable of being decorated by gold to produce the observed substitutional gold profile.

C= A defect with diffusivity D_C which obeys the diffusion equation (5) and can be annihilated at the silicon surfaces.

$$\frac{\partial [C]}{\partial T} = D_C \frac{\partial^2 [C]}{\partial x^2} + \frac{\partial [B]}{\partial T} \qquad (5)$$

Solving equation (5) for [B] using equation (4) and assuming that [A] is constant and that the equilibrium value of k[A] is negligible compared to $[B]^2$ yields the equation:

$$\frac{\partial [B]}{\partial T} = \frac{\partial}{\partial x}\left(\frac{D_C^*}{[B]^2}\frac{\partial [B]}{\partial x}\right) \qquad (6)$$

Where $D_C^* = [C]^{eq}\,[B]^{eq}\,D_C$.

For $A = Au_i$, $B = Au_{si}$, and $C = Si_i$, this equation is the differential equation derived from the KO model, the solution of which provides an excellent fit to experimental data. However, any other model with defects that satisfy the definitions of A, B, and C would also produce a defect profile which matches the experimental profiles for Au_{si}. In particular, we propose a "vacancy cluster" model with $A = V^N$, B = V, and $C = V^{N-1}$ where V^N is a cluster or void containing N vacancies. These vacancy clusters are formed from the high concentration of vacancies quenched into the silicon during crystal growth. We will show experimental results which suggest that the surface of the growing crystal is a rather ineffective sink for the grown-in vacancies. For the case N = 3 we have:

$$V^3 \rightleftharpoons V + V^2$$

This reaction with B = V creates a monovacancy distribution which matches the experimentally obtained U-shaped profile for Au_{si}. As the sample is cooled to room temperature, the supersaturated Au_i atoms would become Au_{si} via the FT mechanism, thereby producing the observed profile.

We note that this model assumes that the Au_i are not active in the formation of the U-shaped vacancy profile. At high temperatures the reverse reaction of (1) is dominant. Only near room temperature, where all experimental measurements of the profile are made, can the forward reaction proceed to completion.

We cannot distinguish between the KO model and the vacancy cluster model by examination of the Au_{si} profile since both models predict the same profile. However, these two models make different predictions when we consider anneals which occur prior to the deposition and diffusion of the Au. The KO model predicts that preanneals occurring before Au diffusion should not have any effect on the Au_{si} profile. The profile predicted by the vacancy cluster model should be affected by the duration of high temperature anneals which occur before the diffusion of the Au since diffusion of monovacancies occur even in the absence of Au. It appears that the effects of preanneals on Au diffusion have not been previously reported. The equipment and experiments designed to test the effects of preanneals on Au diffusion are described below.

EXPERIMENT

The equipment used in these investigations has been described elsewhere [7], the following is a brief description included for

completeness. Double side polished float-zoned 100mm diameter boron-doped silicon wafers of <100> orientation, resistivity of 60-70 ohm-cm, were cleaved into strips prior to processing. The side to be coated with Au was vapor stripped with HF.

Figure 1 shows the apparatus used for the sample annealing and the gold deposition. An ac voltage controlled by an autotransformer was applied to both the sample and heater strip to provide resistance heating. The sample was radiatively heated by the heater strip, a 0.02 ohm-cm boron doped silicon strip mounted 1 mm from the sample, until the sample conductivity was sufficient to permit resistance heating. An optical pyrometer mounted outside the vacuum chamber measures the temperature of the heavily doped strip and controls the autotransformer. The sample is mounted in a vacuum chamber above a resistance heated molybdenum boat containing the gold to be evaporated onto the sample. This equipment allows gold deposition onto the heated sample at any time during the annealing while maintaining the sample at processing temperature.

After processing, the samples were beveled and measured by a spreading resistance probe (SRP). We used a degeneracy factor of 28 [8] to facilitate comparison of the Au profiles presented here with the profiles published by Coffa et al [3].

Figure 2 shows the Au_{Si} profile obtained from a sample which was rapidly heated in vacuum to 1233 K and immediately coated with Au on the side at Depth = 0, and then annealed for 5 minutes at 1233 K to diffuse the Au. The profile for this sample is much more asymmetric than previously published profiles [1,2,3]. Both profiles in this figure were normalized with a Au_{Si}^{eq} value of 3.25E15/cm^3, a value selected to provide the best fit to the calculated curve. This value is within the range of values for Au_{Si}^{eq} given by Coffa [3]. The solid line on the graph is the numerical solution of equation (6) using the published value [3] of D^{*}=1.6E-8 cm^2/sec.

The second sample shown in figure 2 was preannealed in vacuum at 1233 K for 90 minutes prior to the deposition of a Au layer and then annealed at 1233 K for 5 minutes to diffuse the Au. In contrast to the non-preannealed sample, the Au_{Si} profile is nearly one-sided. Despite the longer total annealing time, this sample has a much lower Au_{Si} concentration on the side which did not receive the Au deposition. An additional sample (not shown) was preannealed for 90 minutes at a pressure of 500 mTorr of air prior to the Au deposition. This sample is very similar in appearance to the vacuum annealed sample, but with a higher Au_{Si} concentration at depth = 500 microns.

We have noted [7] similar nearly one-sided profiles under a range of conditions including cases where the sample was annealed for 30 minutes following the Au deposition.

DISCUSSION

For either the FT, KO or vacancy cluster mechanisms, the two-sided profile is caused by the wafer surfaces acting as sources or sinks for the point defects which mediate the diffusion of the Au_{Si}. The one-sided profiles obtained when the diffusion occurs in a vacuum indicate that undamaged silicon surfaces which have been annealed at high temperature consist primarily of steps, lacking energetically costly kink sites and are ineffective as sources or sinks for point defects. The deposition of Au and the subsequent formation of a Au eutectic layer creates the required kink sites

only on the side receiving the Au deposition, resulting in the observed one-sided Au_{Si} profiles. In all the previously reported work on Au diffusion the samples were coated with Au prior to any high temperature processing. We propose that in such samples Au can diffuse to the uncoated side of the sample and stabilize the kink sites before they are annealed out. The sample which did not receive a preanneal shows an asymmetric two-sided profile because the kink sites did not have time to completely disappear prior to arrival of Au_{Si}. This theory is supported by the work of Swartzentruber et al [9] who used an STM to determine step and kink energies on <100> silicon surfaces and noted that their analysis was "complicated by the fact that there are so few kinks."

The inability of an annealed silicon surface to act as a source or sink for point defects supports the proposal that silicon is highly supersaturated with vacancies which originated during crystal growth. These vacancies would be unable to annihilate on the solidified surfaces of the crystal and could only establish equilibrium by diffusing to the solid/liquid interface.

ACKNOWLEDGEMENTS

This work was supported in part by U.S. AFOSR-89-0309.

REFERENCES

1. N.A. Stolwijk, B. Schuster and J. Holzl, Appl. Phys., **A33**, 133 (1984)
2. N.A. Stolwijk, J. Holzl and W. Frank, Appl. Phys., **A39**, 37 (1986)
3. S. Coffa, L. Calcagno, S.U. Campisano, G. Calleri and G. Ferla, J. Appl. Phys., **64**, 6291 (1988)
4. U. Gosele, W. Frank and A. Seeger, Appl. Phys., **23**, 361 (1980)
5. J.A. Van Vechten, U. Schmid, and Zhang Q.-S., J. Electron. Materials, **20**, 431 (1991)
6. W.R. Wilcox and T.J. LaChapelle, J. Appl. Phys., **35**, 240 (1964)
7. R. K. Graupner, J.A. Van Vechten and P. Harwood, J. Vac. Sci. Tech. B, to be published
8. S. Coffa, G. Calleri, L. Calcagno, S.U. Campisano and G. Ferla, Appl. Phys. Lett., **52**, 558 (1988)
9. B.S. Swartzentruber, Y.-W. Mo, R. Kariotis, M.G. Lagally and M.B. Webb, Phys. Rev. Lett., **65**, 1913 (1990)

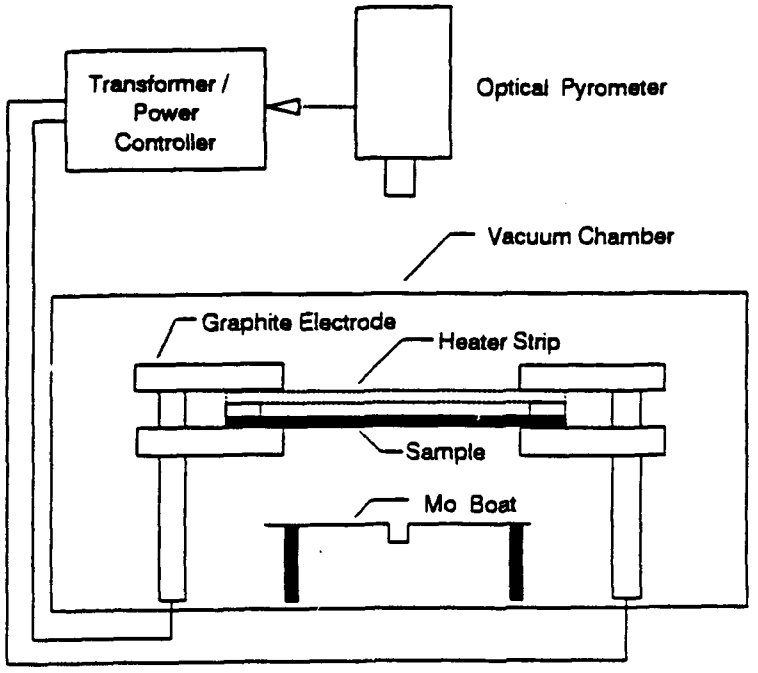


Fig 1. Schematic of the apparatus used for Au deposition and annealing

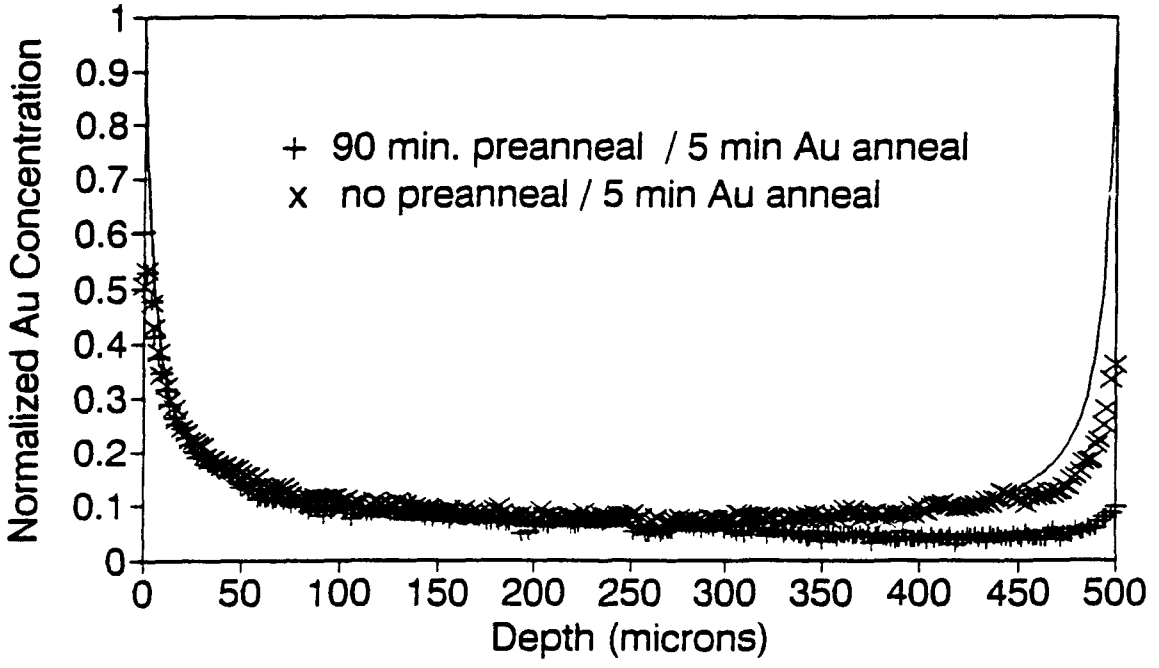


Fig 2. Au concentration profiles for samples with and without a preanneal at 1233 K followed by Au diffusion at 1233 K.

IMPROVED STABILITY OF ZnS:Mn ACTFEL DEVICES

Ron Khormaei,[1] John F. Wager,[1] and Christopher N. King[2]

[1]Department of Electrical & Computer Engineering
Center for Advanced Materials Research
Oregon State University
Corvallis, OR 97331

[2]Planar Systems, Inc.
Beaverton, OR

Abstract

Several approaches have been employed to reduce brightness-voltage instabilities of ZnS:Mn ACTFEL displays. One approach involves the addition of a CaS layer at one or both of the phosphor/insulator interfaces. Another approach involves oxygen exposure of ZnS prior to second insulator deposition. Major improvements are observed, as determined by brightness-voltage aging and latent image experiments.

Introduction

One of the current challenges facing ZnS:Mn alternating current thin film electroluminescent (ACTFEL) device technology is fabrication of displays with variable brightness intensity. This goal, in principle, can be accomplished by varying the driver voltage amplitude; in practice, instabilities in the brightness-voltage (BV) characteristics can lead to luminance shifts. It has been shown that symmetric drive addressing circuitry stabilizes the threshold and saturation luminance.[1] However, there is still the question of intermediate brightness level instability. Our goal for the work described herein was to improve the intermediate BV stability of ZnS:Mn devices.

We have employed two approaches which result in improved ACTFEL device stability. The first technique is to deposit an intermediate CaS layer between the ZnS phosphor and the silicon oxynitride, SiON, insulator. We find, unlike the previous reports,[2] this approach leads to improved stability regardless of whether the CaS is deposited at the top or bottom interface or at both interfaces. A second process modification which yields improved device stability is to expose the ZnS:Mn surface to an oxidizing ambient prior to deposition of the second insulator. We have found that improvement in the device stability is rather insensitive to how the oxygen exposure is accomplished.

Experimental Methods

Device Structures

The electroluminescence displays, for this experiment, are fabricated with a 4-inch by 8-inch active area, 256x512 pixels. Each pixel for a standard panel has a structure as shown in Figure 1 where evaporated ZnS:Mn is the active phosphor layer which is sandwiched between two sputtered SiON insulator layers. Aluminum and ITO electrodes are patterned using standard photolithography methods. Annealing is usually performed in a nitrogen ambient after deposition of the second insulator.

Two device structures are used for improved stability. The first method involves the addition of CaS buffer layers which were added to both or either side

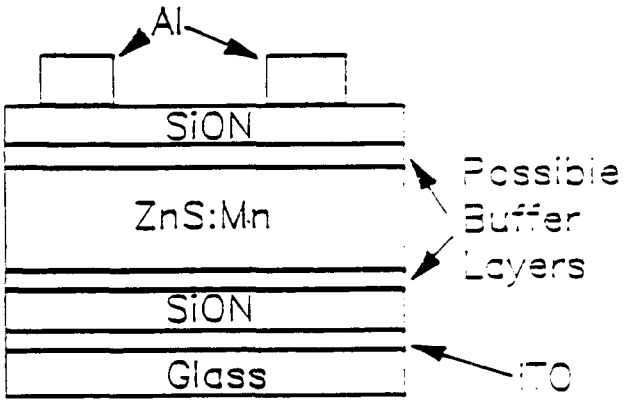


Figure 1. The Electroluminescent Device Structure.

of the ZnS layer (Figure 1). The CaS was deposited by electron beam evaporation. In addition to the variation in the location of the buffer layer, different CaS deposition temperatures were also tested.

The second method involves a modification of the annealing treatment and subsequent oxygen exposure. Rather than annealing in nitrogen after the second insulator deposition, the device is annealed after ZnS deposition but prior to deposition of the second insulator. This anneal is accomplished in vacuum at 460°C. After vacuum annealing, the ZnS surface was exposed to oxygen by heating the sample in air or by oxygen plasma exposure at room temperature. Our experience indicates that the method of oxygen exposure is not significant, whereas a vacuum anneal prior to oxygen exposure is required for improved stability.

Measurement Techniques

Two types of measurements are used to assess the relative device stability. The first test is denoted BVB for brightness voltage bands. A selected set of columns in the panel are driven through an array of resistors for different aging times. This experiment uses drivers which output uniform, symmetric, 200 Hz voltage waveforms simultaneously to all "on" pixels. This measurement technique allows operation of different adjacent columns for different aging times but measurement of all columns at the same time; thus, avoiding problems with drift of the measurement equipment.

The second test employed for assessment of relative device stability is the latent image test. Portions of a full-sized panel are operated continuously for an extended period of time (700 hours for the work reported herein) using commercial, symmetric, 60 Hz drivers while other portions of the panel are left off. After 700 hours of operation, the entire

panel is momentarily turned on and the differential brightness of the aged versus unaged portions of the panel are measured. This differential brightness is then plotted as a function of the brightness of the aged portion of the device, which varies as the peak applied voltage.

Results

BV curves as a function of aging (from the BVB test) are shown in Figures 2 and 3 for, respectively, a standard panel and for a panel with a 100 nm CaS buffer layer inserted at the second ZnS/insulator interface. The improved device stability is evident. We find a similar improvement in the device stability if CaS is present at the first ZnS/insulator interface or if it is present at both interfaces. Thus, the presence of a CaS buffer layer improves the ACTFEL device stability. We find the location of the buffer layer to be unimportant in contrast to that found by investigators at Matsushita, who concluded that improvement could only be obtained when CaS was placed at both interfaces.[2] Thus, the instability mechanism proposed by Matsushita, ion migration from the insulator, does not seem to be operative in our devices.

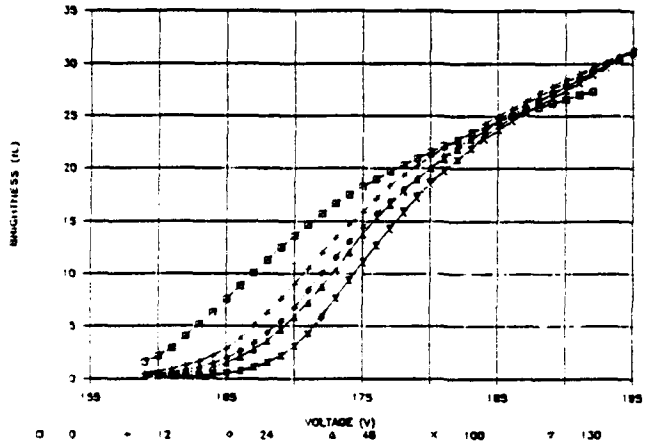


Figure 2. Brightness-Voltage Characteristics as a Function of Aging of a Standard Control Panel.

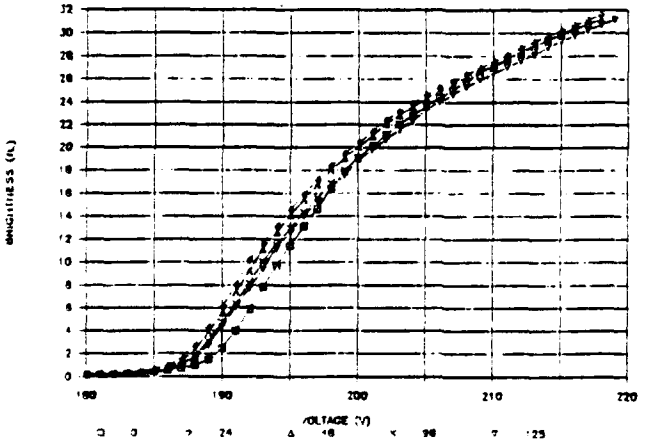


Figure 3. Brightness-Voltage Characteristics as a Function of Aging of a Device with a CaS Buffer Layer at the Second ZnS/Insulator Interface.

The sensitivity of the device stability is a function of the CaS deposition temperature. Three CaS deposition temperatures were investigated: 180, 230, and 320°C (Figure 3 is for 230°C). A slight improvement was observed as the deposition temperature increased. Thus, a higher CaS deposition temperature gives improved device stability.

The BVB test was also applied to the structure with ZnS exposed to oxygen. The BV curves of a panel heated in air after the vacuum anneal is shown in Figure 4. The improved stability compared to the control panel of Figure 2 is clearly evident. Several other panels were prepared using other oxygen exposure procedures; heat treatment in air for temperatures between 150 and 260°C for various oxidation times and oxygen plasma exposure at room temperature for 1-10 min. For all conditions investigated, the improved stability is similar to that shown in Figure 4.

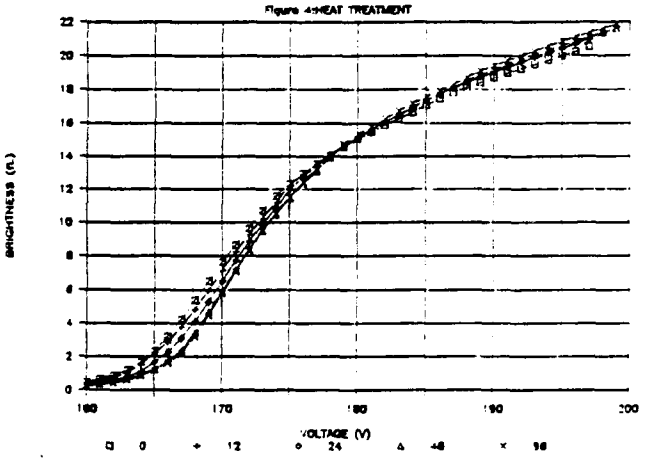


Figure 4. Brightness-Voltage Characteristics as a Function of Aging for an Oxygen-Exposed Panel.

Certain trends in aging are evident from a comparison of Figures 2, 3, and 4. The threshold voltage of the BVB curve increases for the control samples; the residual threshold voltage shift in the BVB plot also increases for the case of oxygen exposed devices whereas it decreases for the case of devices with a CaS buffer layer. The maximum brightness levels are comparable between standard panels and the ones with a CaS buffer layer. However, for the same layer thickness, the brightness level of oxygen-exposed panels is lower than that of a standard panel.

The second type of measurement used to characterize the device stability is the latent image test. The latent image characteristics for one control device, one device with CaS at the second ZnS interface, and one device vacuum annealed and heated in air are shown in Figure 5 for an aging time of 700 hours. Note that an ideal plot would have zero differential brightness for all brightnesses of the aged portion of the device. The differential brightness for this plot is obtained where the brightness of the aged area is below 4 FL, where experience has shown the worst aging effects. The result indicates that, in agreement with BVB results, the two new process modifications actually improve the performance of panels in standard drivers. In general, the latent image test shows the same trends in improved stability as the BVB test. One exception is when the CaS is deposited at the bottom interface; in which case no improvement in stability is observed with the latent image test

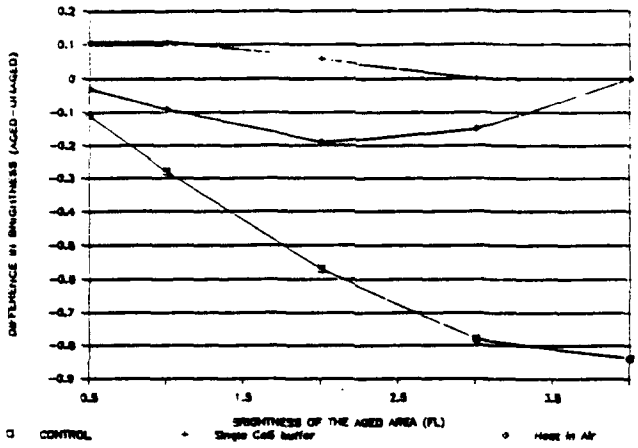


Figure 5. Latent Image Characteristics for a Control, CaS Buffer Layer, and Oxygen Exposed Panels.

compared to the control sample. The latent image test is considered less conclusive than the BVB test because of uncertainties in the driver circuitry and in the sample history; this issue is presently under further consideration.

Conclusions

Two different approaches have been demonstrated to yield improved device stability as monitored by BVB and latent image tests. One approach is to insert a thin buffer layer of CaS at one or both ZnS/insulator interfaces. The second approach involves oxygen exposure of the ZnS surface prior to deposition of the second insulator. The second approach suffers from the fact that the device brightness is significantly reduced by oxygen exposure.

An atomistic explanation for the observed device improvement is currently unclear. We postulate the improvement is associated with a reduction in the number or a reduced tendency of sulfur vacancies to diffuse during device operation. We are currently undertaking experiments to quantify and further test this hypothesis.

References

1. R.T. Flegal and G.N. King, *SID 86 Digest*, 177 (1986).

2. M. Nishikawa, et al., *SID 88 Digest*, 19 (1988).

3. S. Watnabe, et al., Paper of Technical Group of IECE Japan, ED65-6, (1985).

4. R. Coovert, Planar Systems, Private communication.

Capacitance-voltage characteristics of alternating-current thin-film electroluminescent devices

R. C. McArthur, J. D. Davidson, and J. F. Wager
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

I. Khormaei and C. N. King
Planar Systems, Inc., Beaverton, Oregon 97006

(Received 5 October 1989; accepted for publication 26 February 1990)

The capacitance-voltage (C-V) technique is proposed as a method for characterization of the electrical properties of alternating-current thin-film electroluminescent (ACTFEL) display devices. Analysis of the C-V and aging characteristics of ZnS:Mn ACTFEL devices indicates that the C-V technique is complementary to the charge-voltage technique in the extraction of device physics information.

The standard technique used to characterize the electrical properties of alternating-current thin-film electroluminescent (ACTFEL) display devices is charge-voltage (Q-V) analysis.[1-8] The purpose of this letter is to propose a complementary approach for electrical characterization of ACTFEL devices, the capacitance-voltage (C-V) technique. Since capacitance is simply the derivative of charge with respect to voltage, the C-V technique can be viewed as an extension of the Q-V technique.

The C-V measurement is accomplished using the circuit shown in Fig. 1. An ac waveform generator with a small duty cycle drives the ACTFEL device and a 0.5–5 kΩ series resistor. For all of the data present herein, the waveform used is generated with a Wavetek model 275 Arbitrary Waveform Generator and is a 1 kHz symmetric waveform with 5 μs rise and fall times and a 35 μs pulse width measured at 50% of the maximum pulse amplitude. The C-V curve is obtained by recognizing that

$$C = \frac{dq}{dv_2} = \frac{dq/dt}{dv_2/dt} = \frac{i}{dv_2/dt}. \tag{1}$$

The current may be obtained as the voltage difference across a series resistor:

$$i = (v_1 - v_2)/R_s. \tag{2}$$

In Eqs. (1) and (2), v_1 and v_2 are voltages with respect to ground as indicated in Fig. 1 and are obtained by sampling with a Tektronix model 7854 digitizing oscilloscope. The mathematical operations defined by Eqs. (1) and (2) are accomplished with the digitizing oscilloscope and the C-V curve is obtained by ploting C vs v_2. The experiment is controlled with a personal computer which is also used for data storage and further data processing, if required. The viability of this method of measuring capacitance was tested by replacing the ACTFEL device with a capacitor of known value; the known capacitance was accurately determined using the above technique.

To understand the nature of the C-V measurement, first consider the ideal Q-V curve of an ACTFEL device, shown in Fig. 2, operating in steady state in a light emission regime. Differentiating Fig. 2 results in the ideal C-V curve illustrated in Fig. 3. The minimum capacitance is that due to the total capacitance of the ACTFEL device (i.e., a series combination of the phosphor and insulator capacitances) while the maximum capacitance is that due exclusively to the insulators as the phosphor is conducting during the charge transfer portion of the waveform.[1-8] Additionally, the threshold voltage indicative of the onset of charge transfer V_{th} and the maximum applied voltage amplitude V_m are indicated in Fig. 3.

In actual measurements, the C-V curve is normally acquired individually for each voltage polarity and only the first half of the driver pulse waveform is acquired. Each voltage polarity waveform is acquired individually because of the extremely small duty cycle of the driver waveform which precludes high-resolution digital acquisition of two subsequent waveforms. Only the first half of the driver pulse waveform is usually acquired since the information of most relevance is contained in this regime. The second half of the pulse is essentially due to the retrace which is manifest in Fig. 3 as the portion of the C-V rectangle indicated by the dashed lines. Thus, an ideal standard C-V curve would look like the right half of the curve shown in Fig. 3 with the dashed line traces absent.

A Q-V curve of a ZnS:Mn ACTFEL device is shown in Fig. 4. The charge Q for this curve is obtained from an integration of the current found in Eq. (2). The slopes of the straight line fits to the Q-V characteristic yield the insulator and total device capacitances, as in the ideal Q-V case. The intersection of these two straight line fits defines the threshold voltage. However, note deviations from ideal linear behavior near V_{th} and V_m.

To investigate these deviations from ideality in more

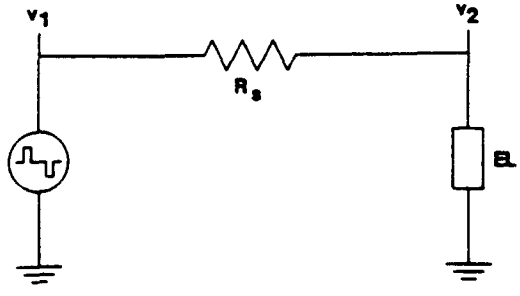


FIG. 1. Circuit used for C-V analysis.

 0003-6951/90/191889-03$02.00

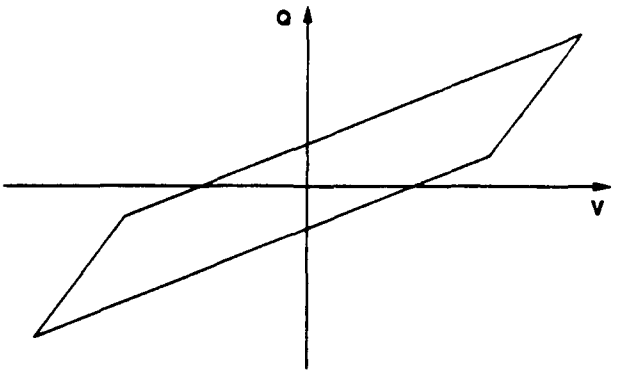


FIG. 2. Ideal Q-V curve.

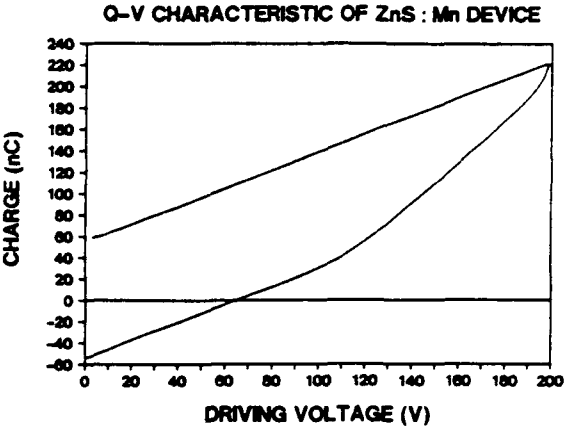


FIG. 4. Q-V curve for a ZnS:Mn ACTFEL device.

detail, it is convenient to focus attention on the C-V curve of Fig. 5 which is derived from the same data as Fig. 4. Before Eq. (1) was used to calculate C, the numerically computed voltage derivative was filtered via least-squares approximation using piecewise, quadratic polynomials with B-spline basis functions. This operation suppresses the amplified noise due to differentiation. The horizontal portions of the C-V curve are due to the constant capacitances associated with the total and insulator capacitances. The transition region near V_{th1} corresponds to the initiation of transferred charge from localized states near the interface across the ZnS, thereby exciting Mn atoms into luminescently active excited states and shunting the ZnS capacitance.

It is evident from Fig. 5 that the threshold voltage is not as clearly defined as implied by the ideal curve of Fig. 3. In fact, three different definitions of the threshold voltage from Fig. 5 are possible corresponding to the onset of the transition, the midpoint of the transition, and the saturation of the transition in which the capacitance is exclusively that of the insulator capacitance, which we denote V_{th1}, V_{th2}, and V_{th3}, respectively. We have found that V_{th1} varies most readily with changes in the pulse width or voltage amplitude of the applied waveform while V_{th3} is relatively insensitive to these variations. We attribute V_{th1} to the onset of emission of electrons from shallow interface traps. Also, we believe that V_{th3} corresponds to the initiation of field clamping. We have also found that V_{th2} corresponds most closely to the threshold voltage found from Q-V measurements.

Total and insulator capacitances calculated on the basis of the thicknesses and dielectric constants of the insulator and phosphor layers are found to be in good agreement with values obtained from the C-V measurement. It is apparent from this discussion that the C-V measurement reveals more device physics information regarding the transferred charge threshold characteristics than the Q-V measurement.

The second deviation from ideality which occurs in Fig. 5 is an increase in the measured capacitance at voltages near V_m. This deviation is also apparent in the Q-V curve of Fig. 4. There are two possible explanations for this increase in capacitance at large applied voltages. First, the insulator leakage is large enough that the appropriate circuit model for the insulator is a resistor in parallel with the insulator capacitance such that the apparent rise in capacitance is due to a RC effect. Second, transferred charge is injected into the insulator such that the effective thickness of the insulator is reduced and the corresponding capacitance of the insulator is increased. Further work is required in order to determine which explanation is preferred.

The C-V technique is useful for aging analysis of ACTFEL devices. The aging characteristics of a ZnS:Mn device are indicated in Fig. 6. The C-V characteristics shift to higher voltages with respect to aging time. Note that this shift in the C-V curves is an essentially rigid shift and is found to be

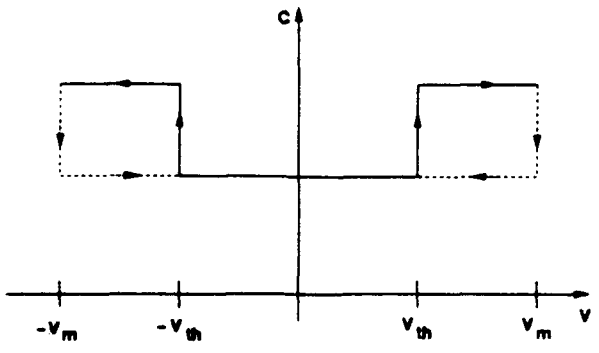


FIG. 3. Ideal C-V curve obtained from differentiation of Fig. 2.

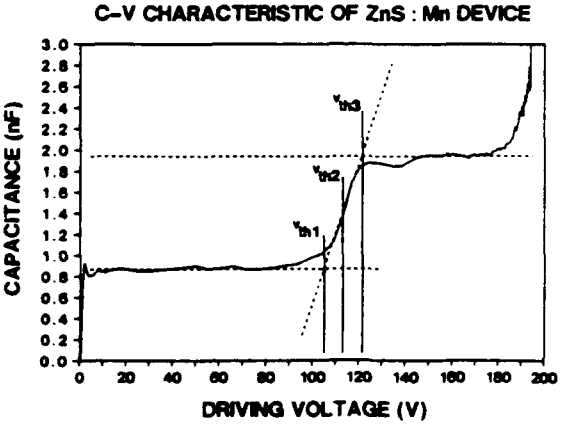


FIG. 5. C-V curve for a ZnS:Mn ACTFEL device.

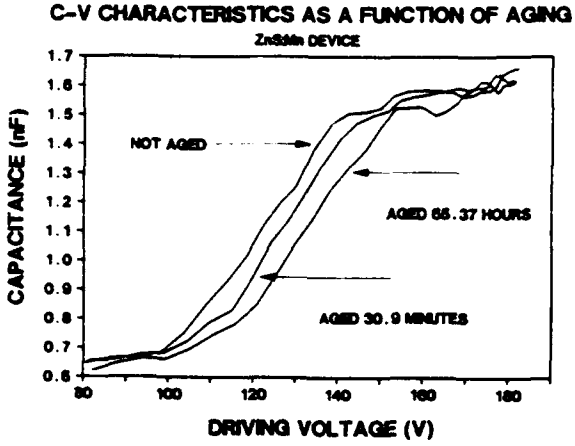


FIG. 6. C-V curves as a function of aging time for a ZnS:Mn ACTFEL device.

logarithmic with respect to aging time. Although not evident from Fig. 6, it is found that the insulator and total capacitances are independent of aging time, except near V_m.

In summary, we propose the C-V technique as a means of exploring the electrical properties of ACTFEL display devices and we demonstrate the utility of the technique from an analysis of the C-V and aging characteristics of a ZnS:Mn device. Although the C-V technique may be regarded as an extension of the Q-V technique, we believe C-V analysis to be more readily interpretable in terms of the physics of the ACTFEL device.

This work was supported in part by the Air Force Office of Scientific Research under contract No. AFOSR 89-0309.

[1]Y. S. Chen and D. C. Krupka, J. Appl. Phys. **43**, 4089 (1972).
[2]D. H. Smith, J. Lumin. **23**, 209 (1981).
[3]K. W. Yang, S. J. T. Owen, and D. H. Smith, IEEE Trans. Electron Devices **ED-28**, 703 (1981).
[4]K. W. Yang, Ph.D. thesis, Oregon State University, 1981.
[5]R. Mach and G. O. Müller, Phys. Status Solidi A **69**, 11 (1982).
[6]P. M. Alt, Proc. SID **25**, 123 (1984).
[7]Y. A. Ono, H. Kawakami, M. Fuyama, and K. Onisawa, Jpn. J. Appl. Phys. **26**, 1482 (1987).
[8]E. Bringuier, J. Appl. Phys. **66**, 1316 (1989).

Ph 170

UDC 535.376:061.3, 539.23:538.975

ACTA POLYTECHNICA SCANDINAVICA

APPLIED PHYSICS SERIES No. 170

5th International Workshop on Electroluminescence

Edited by
MARKKU LESKELÄ and
ERJA NYKÄNEN

HELSINKI 1990

CHARACTERIZATION OF ZnS:Mn AC THIN-FILM ELECTROLUMINESCENT DEVICES BY CAPACITANCE-VOLTAGE ANALYSIS

R.C. McArthur,* J.D. Davidson,* I. Khormaei,† J.F Wager,* and C.N. King†
*Department of Electrical & Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, OR 97331-3211, USA.
†Planar Systems, Inc., Beaverton, OR 97006, USA.

ABSTRACT

A new method for characterizing the electrical properties of alternating-current thin-film electroluminescent (ACTFEL) devices, the capacitance-voltage (C-V) technique, is described. The C-V technique is complementary to the charge-voltage (Q-V) technique which is conventionally employed for ACTFEL device analysis.

1. THE CAPACITANCE-VOLTAGE (C-V) TECHNIQUE

The C-V measurement[1] is accomplished using an arbitrary waveform generator (Wavetek Model 275) with a small duty cycle to drive the ACTFEL device and a 1.5 k Ω series resistor. The standard waveform used is symmetric with a frequency of 1 kHz, rise and fall times of 5 μs, and a pulse width of 30 μs where the pulse width is defined as the time over which the waveform is constant at its maximum voltage. The waveform voltage, v_1, and the voltage across the ACTFEL device, v_2, are monitored with a digitizing oscilloscope (Tektronix Model 7854) as indicated in Fig. 1. The C-V curve can then be obtained by recognizing that

$$C = \frac{dq}{dV_2} = \frac{dq/dt}{dv_2/dt} = \frac{i}{dv_2/dt} . \qquad (1)$$

The current is obtained as the voltage difference across the series resistor:

$$i = (v_1 - v_2)/R_s . \qquad (2)$$

The mathematical operations defined by Eqs. (1) and (2) are accomplished with the digitizing oscilloscope and the C-V characteristics are obtained by plotting C versus v_2. The experiment is controlled with a personal computer which is also used for data storage and further data processing, if required.

2. INFORMATION AVAILABLE FROM C-V CURVES

A C-V curve for a ZnS:Mn ACTFEL device is illustrated in Fig. 2. The minimum capacitance at low voltages is due to the total capacitance of the ACTFEL device (i.e. a series combination of the phosphor and insulator capacitances). The constant capacitance at larger applied voltages, indicated by the upper dashed line, corresponds to the insulator capacitance only, as the phosphor is conducting during portion this of the waveform, effectively shunting the phosphor capacitance. The structure in the C-V curve near saturation (i.e. in the voltage

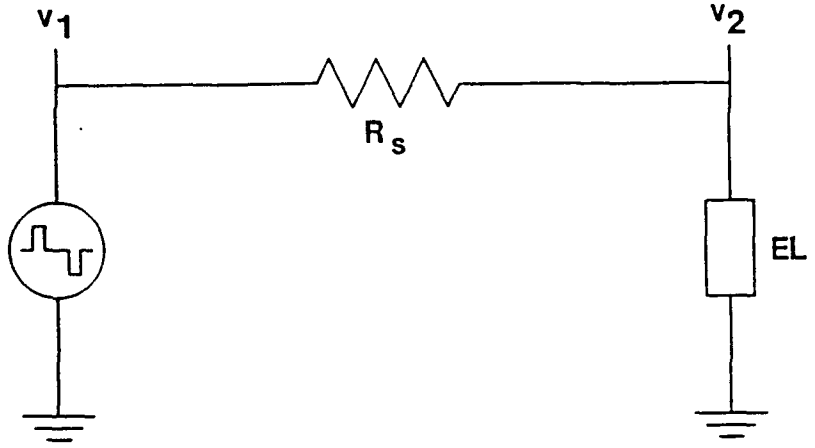


Figure 1. The circuit used for C-V analysis.

range of 120-140 V) is not always observed; we have not yet been able to correlate such structure with the materials physics of the device. We currently believe the capacitance rise near the maximum voltage to be an artifact of the measurement.

Note from Fig. 2 that the threshold voltage is not uniquely defined. We define three threshold voltages corresponding to the onset of the C-V transition, the midpoint of the transition, and saturation of the transition, which we denote as V_{th1}, V_{th2}, and V_{th3}, respectively. We attribute V_{th1} to the onset of the emission of electrons from shallow interface traps and V_{th3} to the initiation of field clamping (i.e. the external voltage at which the interface trap density is of sufficient magnitude such that the ZnS field remains constant). V_{th2} corresponds most closely to the threshold voltage found from Q-V measurements. The slope of the C-V curve in the transition region is related to the total density of interface states in the subthreshold region.[2]

3. C-V CURVES OF SYMMETRIC DRIVING WAVEFORMS

C-V analysis of ZnS:Mn ACTFEL devices as a function of maximum voltage amplitude, V_m, pulse width, PW, and rise time, τ, is accomplished using symmetric driving voltage waveforms. The purpose of this set of experiments is to establish the sensitivity of the threshold voltage as a function of the waveform parameters. Care is taken to sufficiently pre-age the sample so that aging effects are minimal. A rigid shift in the C-V curve to smaller thresholds is observed as V_m is increased. The shift in V_{th} is linear with respect to V_m for voltages up to approximately 200 V, above which V_{th} depends sublinearly on V_m as if an onset of saturation is being approached. The C-V curve also shifts rigidly as a function of PW. V_{th} decreases with increasing PW for times up to about 100 μs, above which an onset of saturation begins in which V_{th} depends only weakly on PW. The C-V curve, and hence V_{th}, is virtually independent of τ.

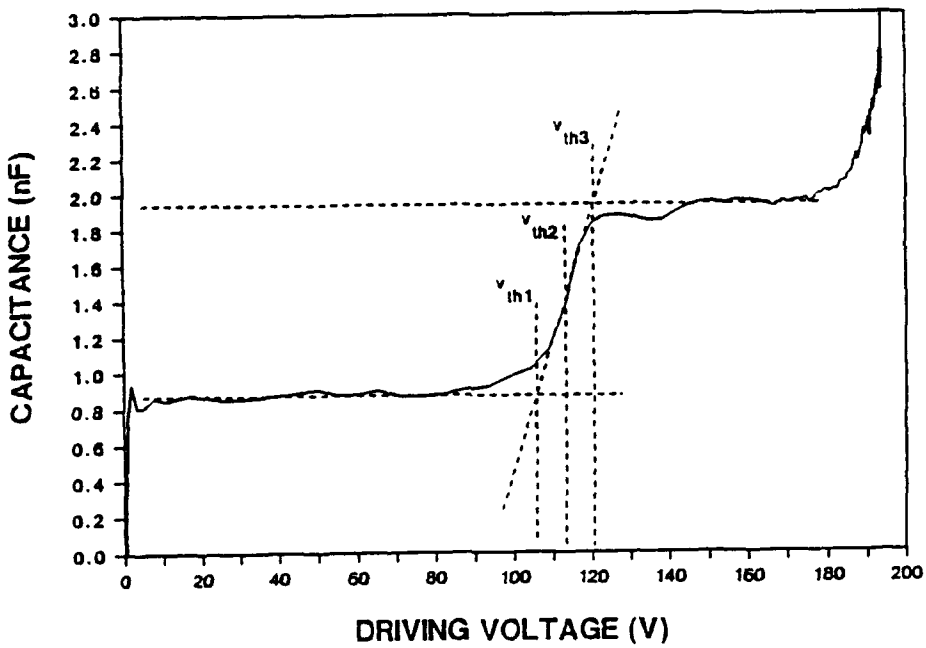


Figure 2. A C-V curve for a ZnS:Mn ACTFEL device

ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Office of Scientific Research under Contract No. AFOSR 89-0309.

REFERENCES

1. McArthur, R.C., Davidson, J.D., Wager, J.F., Khormaei, I., and King, C.N., Capacitance-Voltage Characteristics of Alternating Current Thin-Film Electroluminescent Devices, Appl. Phys. Lett. 56 (1990) 1889-1891.
2. Williams, R. (unpublished).

Ph 170

UDC 535.376:061.3, 539.23:538.975

ACTA POLYTECHNICA SCANDINAVICA

APPLIED PHYSICS SERIES No. 170

5th International Workshop on Electroluminescence

Edited by
MARKKU LESKELÄ and
ERJA NYKÄNEN

HELSINKI 1990

AGING INSTABILITIES OF ZnS:Mn AC THIN-FILM ELECTROLUMINESCENT DEVICES

J.D. Davidson,* J.F. Wager,* I. Khormaei,† and C.N. King†
*Department of Electrical & Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, OR 97331-3211, USA.
†Planar Systems, Inc., Beaverton, OR 97006, USA.

ABSTRACT

Aging characteristics of ZnS:Mn alternating-current thin-film electroluminescent (ACTFEL) devices are monitored via the capacitance-voltage (C-V) technique. Arrhenius analysis of variable temperature aging data indicates the existence of two aging mechanisms, both displaying logarithmic kinetics. The activation energies of these two mechanisms are consistent with electron injection and trapping in the insulator near the interface for room temperature and below and with atomic migration for above room temperature.

1. ACTFEL AGING EXPERIMENTS

ACTFEL aging experiments are accomplished by measuring the threshold voltage, as monitored via the C-V technique,[1,2] as a function of aging time at a constant temperature. The C-V curve is observed to move monotonically to larger voltages as a function of aging time. The slope of the C-V curve decreases slightly as a function of aging which implies that the density of interface traps in the subthreshold region increases with aging time. The threshold voltage shift is well described by logarithmic kinetics, $V_{th} \propto \ln t$, except during the initial aging period. We attribute the deviation from logarithmic kinetics as due to the preaging warm-up procedure. The aging kinetics as a function of temperature are shown in Fig. 1.

An Arrhenius analysis of the data shown in Fig. 1 is accomplished by ignoring the initial data point for each data set (as this data point is attributed to the warm-up procedure), performing linear regression on each dataset, plotting ΔV_{th} as a function of time from the linear regression curves, and undertaking an Arrhenius analysis of the resulting ΔV_{th} versus ln t curves. The resulting Arrhenius plot is indicated in Fig. 2.

2. DISCUSSION OF ACTFEL AGING MECHANISMS

From an analysis of Fig. 1, it is evident that at zero aging, V_{th} is temperature-dependent and increases with increasing temperature. This is in contrast to results of previous researchers[3,4] who report that V_{th} is essentially temperature-independent. Also note from Fig. 1 that the slope of the kinetic curves decreases with increasing temperature and shows evidence of the approach of saturation at high temperature.

Two distinct aging mechanisms are distinguished by their respective activation energies as shown in Fig. 2. We note that a weak temperature dependence (e.g. E_a = 0.067 eV) is expected for a mechanism based on electron injection and trapping at the interface.[5-8] In contrast, a mechanism involving atomic migration (e.g. sulfur-vacancy migration[9]) would be a more likely identification of the high-temperature aging mechanism with an

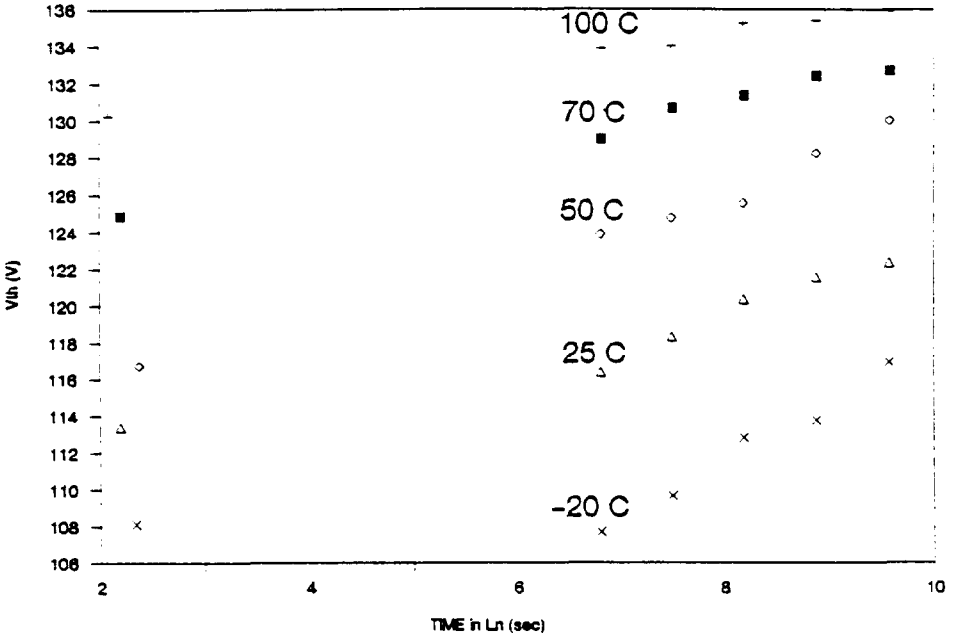


Figure 1. Threshold voltage versus aging time as a function of temperature.

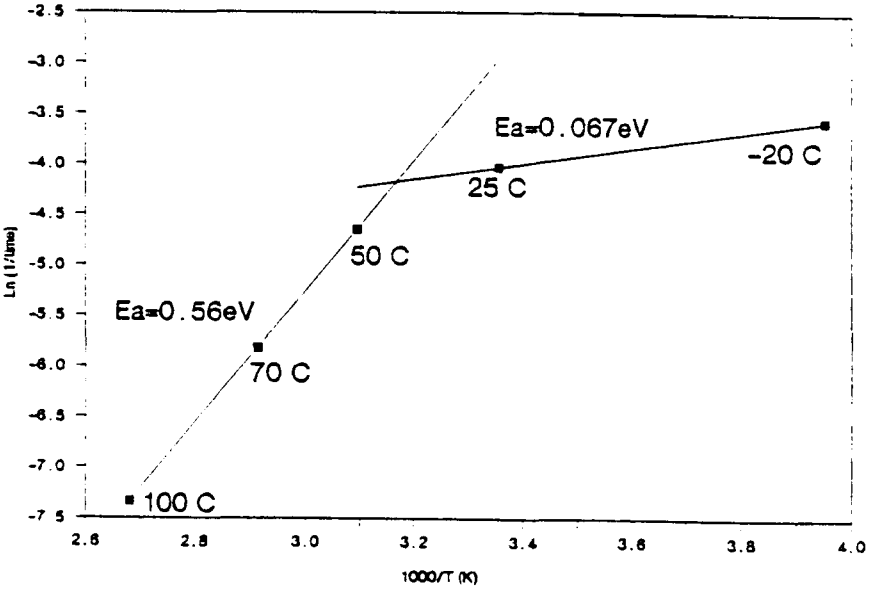


Figure 2. Arrhenius plot of data from Fig. 1.

activation energy of 0.56 eV.[8] Since the active region of the ACTFEL display almost certainly operates at a temperature in excess of 40°C, we conclude from Fig. 2 that the dominant aging instability mechanism in the ZnS:Mn devices we have characterized can be ascribed to atomic migration near the ZnS/insulator interface. We postulate this atomic migration to be due to the migration of sulfur vacancies.[9]

ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Office of Scientific Research under Contract No. AFOSR 89-0309.

REFERENCES

1. McArthur, R.C., Davidson, J.D., Wager, J.F., Khormaei, I., and King, C.N., Capacitance-Voltage Characteristics of Alternating Current Thin-Film Electroluminescent Devices, Appl. Phys. Lett. 56 (1990) 1889-1891.
2. McArthur, R.C., Davidson, J.D., Wager, J.F., Khormaei, I., and King, C.N., Characterization of ZnS:Mn AC Thin-Film Electroluminescent Devices by Capacitance-Voltage Analysis (These Proceedings).
3. Chen, Y.S. and Krupka, D.C., Limitation Imposed by Field Clamping on the Efficiency of High-Field AC Electroluminescence in Thin Films," J. Appl. Phys. 43 (1972) 4089-4096.
4. Yang, K.W., Owen, S.J.T., and Smith, D.H., Studies of Temperature Effects in AC Thin-Film EL Devices, IEEE Trans. Electron Devices ED-28 (1981) 703-708.
5. Chang, J.J., Theory of MNOS Memory Transistor, IEEE Trans. Electron Devices ED-24 (1977) 511-518.
6. Walden, R.H., A Method for the Determination of High-Field Conduction Laws in Insulating Films in the Presence of Charge Trapping, J. Appl. Phys. 43 (1972) 1178-1186.
7. Powell, M.J., van Berkel, C., and Hughes, J.R., Time and Temperature Dependence of Instability Mechanisms in Amorphous Silicon Thin-Film Transistors, Appl. Phys. Lett. 54 (1989) 1323-1325.
8. Juang, M.T., Wager, J.F., and Van Vechten, J.A., Phosphorous Vacancy Nearest Neighbor Hopping Induced Instabilities in InP Capacitors, J. Electrochem. Soc. 135 (1988), 2019-2027.
9. Khormaei, R., Wager, J.F., and King, C.N., Improved Stability of ZnS:Mn ACTFEL Devices, SID 89 Digest (1989) 65-67.

6.5: Electrical Characterization and SPICE Modeling of ZnS:Mn ACTFEL Devices

James D. Davidson, John F. Wager

Oregon State University, Corvallis, OR

Ron Khormaei

Planar Systems, Inc., Beaverton, OR

Introduction

Electrical characterization of alternating-current thin-film electroluminescent (ACTFEL) devices is usually accomplished by charge-voltage (Q-V) [1] or current-voltage (I-V) [2] analysis. We have recently described the capacitance-voltage (C-V) technique [3] as an alternative method for ACTFEL electrical characterization. Our goal for the work described herein is to show that the C-V technique, in conjunction with SPICE modeling using a simple equivalent circuit, provides a powerful means of probing ACTFEL device physics.

The Capacitance-Voltage Technique

C-V analysis is accomplished using the circuit shown in Fig. 1. An ac waveform generator (Wavetek model 275) with a small duty cycle is used to drive a series resistor, R_s, the ACTFEL device, and a current sense resistor, R_i. R_s is chosen to be large compared to R_i (typically $R_s \approx 1.5$ kΩ and $R_i \approx 10$ Ω) so that the voltage drop across R_i is negligible. The three voltages indicated in Fig. 1, v_1, v_2, and v_3, correspond to the driver, the ACTFEL device, and the current sense voltages, respectively.

The standard waveform employed in this work is symmetric with bipolar pulses of trapezoidal shape with 5 μs rise and fall times and a pulse width of 30 μs where the pulse width is defined as the duration in which the pulse is at its maximum amplitude. The frequency of the waveform is 1 kHz.

Voltages $v_1(t)$, $v_2(t)$, or $v_3(t)$ are obtained by sampling using a Tektronix model 7854 digitizing oscilloscope. An oscilloscope time base of 2 μs/division and 128 points per waveform gave best results.

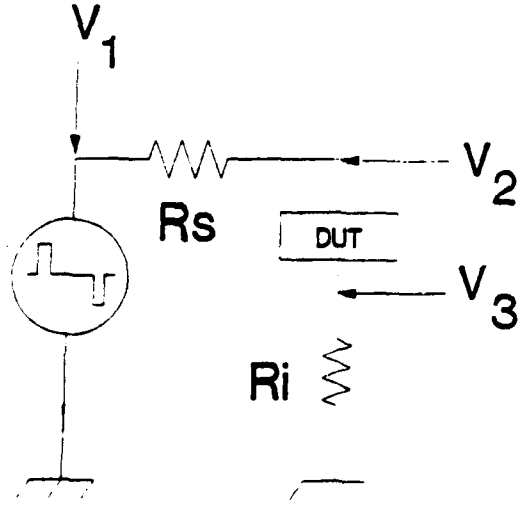


Figure 1. Circuit used for ACTFEL electrical characterization.

For the work described in this paper, only $v_2(t)$ and $v_3(t)$ are normally measured.

Using the circuit shown in Fig. 1, the current is obtained from the voltage across the sense resistor,

$$I(t) = \frac{v_3(t)}{R_i} . \qquad (1)$$

The capacitance is equal to the current divided by the derivative of the voltage across the ACTFEL device [3],

$$C(t) = \frac{I(t)}{d[v_2(t) - v_3(t)]/dt} . \qquad (2)$$

and the C-V curve is then obtained by plotting C(t) versus $[v_2(t) - v_3(t)]$.

The SPICE Equivalent Circuit

The equivalent circuit for SPICE simulation of the ACTFEL device is shown in Fig. 2. C_{i1} and C_{i2} are the insulator capacitances whereas C_p is the phosphor capacitance. The phosphor capacitance is shunted by two back-to-back Zener diodes which are used to account for conduction and field clamping in the phosphor region. These circuit components constitute what we term an ideal circuit model for the ACTFEL device.

Five additional resistors are added to the ideal equivalent circuit to obtain the SPICE equivalent circuit of Fig. 2. R_{i1}, R_{i2}, and R_p represent the parallel resistance of the respective insulator or phosphor layers which would be associated with dc leak-

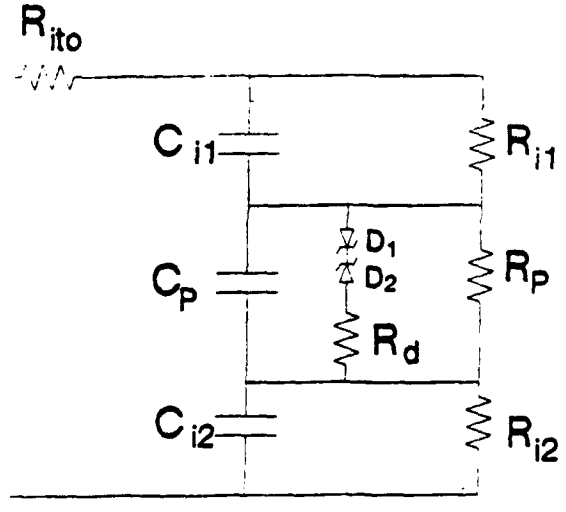


Figure 2. Equivalent circuit for SPICE simulation of ACTFEL devices.

ISSN0097-0966X/91/0000-077-$1.00 + .00 © 1991 SID

age through these layers. R_{ito} is a lumped resistance accounting for the non-zero series resistance of the ITO electrode. R_d we denote as the diode resistance.

SPICE Simulation of C-V Curves

A comparison of a typical C-V curve with a SPICE-simulated C-V curve is shown in Fig. 3. The agreement between the experimental and simulated curve is reasonably good except near the maximum applied voltage. This region of the C-V curve is very noisy because the capacitance arises from the ratio of the measured current and the derivative of the voltage across the ACTFEL device; these signals are both very small near the maximum voltage which results in a small signal-to-noise ratio.

To explore trends in the the parasitic resistance, C-V curves with parametric variations of R_{ito}, R_{i1}, R_p, and R_d around nominal parameter values are shown in Figs. 4-7. Nominal circuit component values used in the SPICE modeling are collected in Table 1. Several trends emerge from an analysis of these figures. First, it is clear that the capacitance rise near zero voltage is

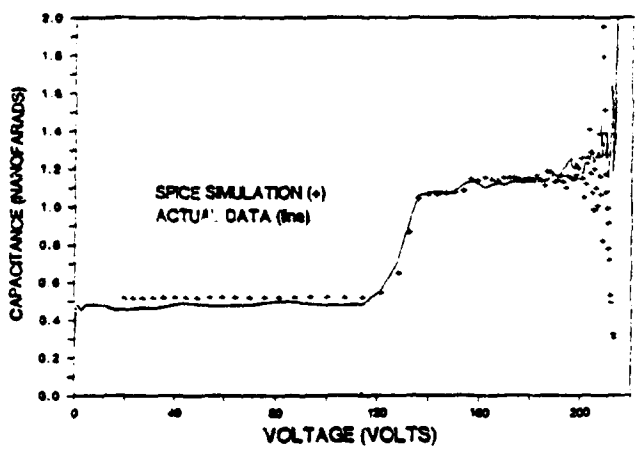


Figure 3. Measured and SPICE-simulated C-V curves.

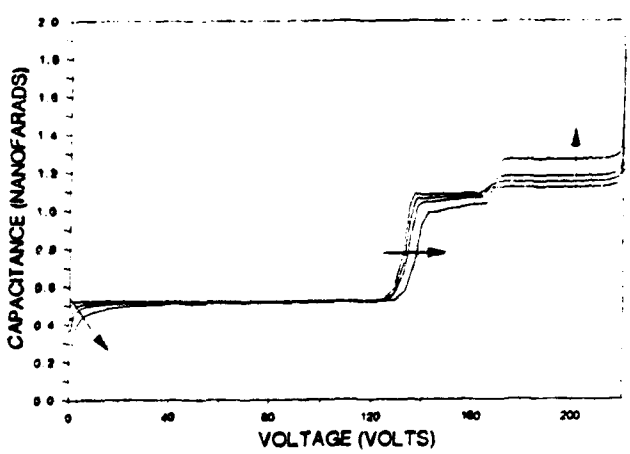


Figure 4. C-V curves for parametric variations of R_{ito}. R_{ito} = 0, 20, 40 and 80 Ω. Arrows indicate an increasing value for R_{ito}.

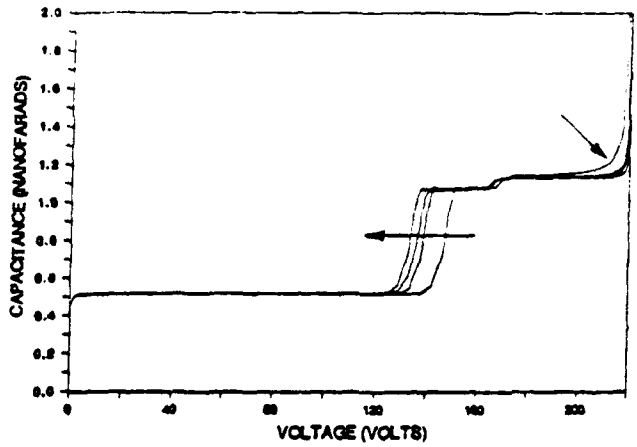


Figure 5. C-V curves for parametric variations of R_{i1}. R_{i1} = 100 k, 500 k, 1 M, and 100 × 10^9 Ω. Arrows indicate an increasing value for R_{i1}.

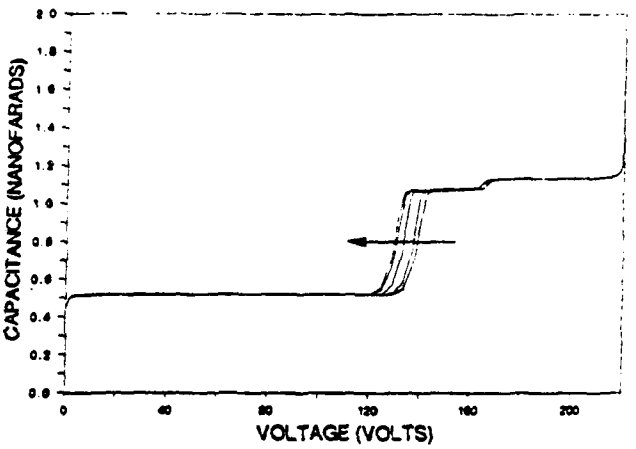


Figure 6. C-V curves for parametric variations of R_p. R_p = 800 k, 1 M, 2 M, 5 M, and 10 M Ω. Arrow indicates an increasing value for R_p.

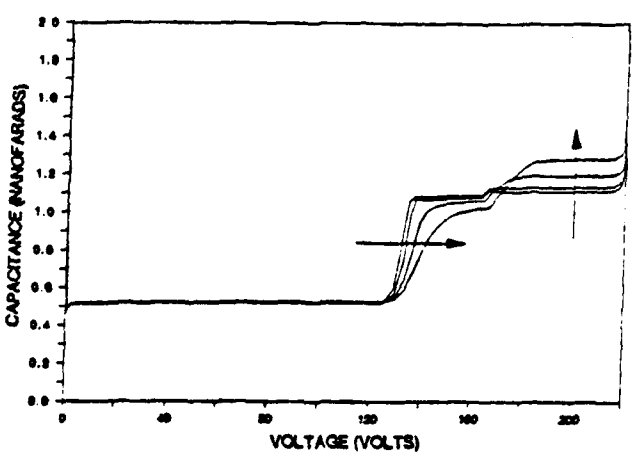


Figure 7. C-V curves for parametric variations of R_d. R_d = 0, 30, 100, and 200 Ω. Arrows indicate an increasing value for R_d.

78 • SID 91 DIGEST

Table 1. Nominal Component Values for the ACTFEL SPICE Model

Component	Component Parameter	Component Value
C_{i1}	Capacitance	2.18 nF
C_{i2}	Capacitance	2.20 nF
C_p	Capacitance	0.987 nF
D_1	Breakdown voltage	92 V
D_2	Breakdown voltage	92 V
R_{ito}	ITO series resistance	20 Ω
R_{i1}, R_{i2}	Insulator parallel resistance	10^{11} Ω
R_p	Phosphor parallel resistance	2×10^6 Ω
R_d	Diode resistance	30 Ω

associated with R_{ito} and that this rise is essentially an RC effect. It is evident from Fig. 4 that R_{ito} influences the abruptness of the C-V transition. R_{ito} is also the most important parameter in determining the capacitance step in the C_i regime of the C-V curve.

From an analysis of Figs. 5 and 6 we conclude that an increase in the insulator or phosphor shunt resistances leads to a rigid shift in the C-V transition to lower threshold voltages. A physical interpretation of this trend is that a reduction in the resistance of the phosphor or insulator layers causes leakage of the polarization charge which gives rise to a concomitant increase in the threshold voltage. Note also that the capacitance rise near V_{th} is affected by a small insulator shunt resistance, as shown in Fig. 5.

As indicated in Fig. 7, R_d affects the abruptness of the C-V threshold with smaller R_d yielding a more abrupt transition. Physically, R_d corresponds to a resistance associated with hot electron injection from interface traps. Thus, R_d is related to the characteristic time for electron emission from interface traps such that we interpret a smearing out of the C-V transition as arising from a higher density of interface traps with characteristic emission times of comparable magnitude to that of the driver waveform (i.e., microseconds). Although the abruptness of the C-V transition is mainly established by the magnitude of R_d, the abruptness is also affected, but to a much lesser extent, by R_{ito}.

Increasing R_d also leads to a larger C_i step in the insulator capacitance regime of the C-V curve. Thus, R_{ito} and R_d are the two most important parameters in determining the C_i step in the insulator capacitance regime. It should be noted that we often, but not always, observe such a C_i step in real C-V curves. Although it is evident that the C_i step depends mainly on R_{ito} and R_d, it is not yet clear why changes in these parameters result in such a step.

Conclusions

In this paper, we have demonstrated the utility of SPICE modeling in conjunction with a simple equivalent circuit model for the ACTFEL device for the interpretation of C-V curves. We believe that such modeling will provide new device physics insight into the nature of ACTFEL operation.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research under Contract No. AFOSR 89-0309.

References

1. D.H. Smith, J. Lumin. 23, 209 (1981).

2. G.O. Müller, R. Mach, B. Selle, and G. Schulz, Phys. Stat. Sol. (a) 110, 657 (1988).

3. R.C. McArthur, J.D. Davidson, J.F. Wager, I. Khormaei, and C.N. King, Appl. Phys. Lett. 56, 1889 (1990).

6.4: Stabilization of ZnS:Mn ACTFEL Devices Through Processing Modifications

Ron Khormaei, Christopher N. King, Richard E. Coovert
Planar Systems, Inc., Beaverton, OR
John F. Wager
Oregon State University, Corvallis, OR

Abstract

Brightness-voltage instabilities of ZnS:Mn ACTFEL devices are significantly reduced through thermal post-deposition methods which increase the sulfur content of the ZnS layer. These improvements support the assertion that the observed instabilities are mainly due to sulfur vacancies. Worst case differential aging after 1000 hours of operation are now well within the requirements of 16-level gray scale monitors.

Introduction

A common challenge for all the flat panel technologies is the stability of display performance with usage. This problem becomes more pronounced when trying to produce gray scale by holding driving parameters at intermediate excitation levels.[1,2] In AC Thin Film Electroluminescent (ACTFEL) devices, it has been shown that symmetric drive circuitry stabilizes the threshold and saturation luminance.[3] The gray scale application has a more severe requirement allowing only minimal changes throughout the intermediate brightness levels. Our goal for this work was to stabilize the B-V curve to a point of easily being acceptable for use in a gray scale monitor. The approach chosen was to establish the mechanism responsible for the improvements observed in previous experiments using CaS buffer layers, [4,5] and to find a more manufacturable method for stabilizing the structure further.

Experimental Procedure and Results

The electroluminescent devices employed in this set of experiments were fabricated with co-evaporated ZnS:Mn sandwiched between two sputtered insulator layers. Based on previous experiments on similar structures, the source of instability was postulated to be sulfur vacancies.[5,6,7] In this work, two post-deposition methods were used to increase the sulfur content of the ZnS film, i.e. reducing the concentration of sulfur vacancies. Both approaches involve thermal treatment of the deposited ZnS film before the second insulator deposition. One method is to anneal the sample in an H_2S atmosphere. The other approach is to heat the sample in an enclosed volume with solid sulfur present. Both methods of sulfur exposure show similar effects in terms of improved device performance.

Two methods are used for evaluating the device stability. The first method is aging and brightness-voltage (B-V) measurements on single test devices. The goal of this type of experiment is to establish the capabilities of the intrinsic device structure without possible complications from matrix arrangements or the drive electronics. This test is performed with high aging frequency of 1000 Hz for accelerated aging and the B-V measurements are performed at 60 Hz after various aging intervals. The resulting curves are plotted by extrapolating the aging period from 1000Hz to 60 Hz (a multiplier factor of 16).

The evaluation of test devices for a traditional structure is shown in Figure 1 and the result after sulfur exposure is shown in Figure 2. As can be seen from Figures 1 and 2, sulfur exposure dramatically increases the stability of the device. The conventional structure shows a threshold voltage (Vth) shift of about 8 volts in the first 160 hours of aging. In the same aging period, the sulfur-exposed samples display a minimal amount of shift.

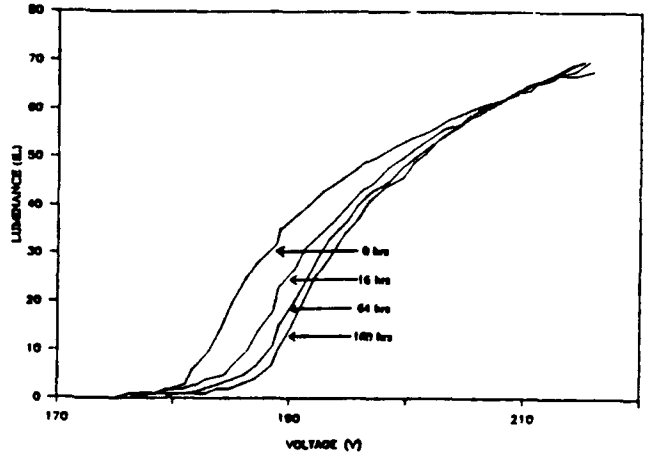


Figure 1. B-V curves for traditional structure - aging time extrapolated to 60 Hz.

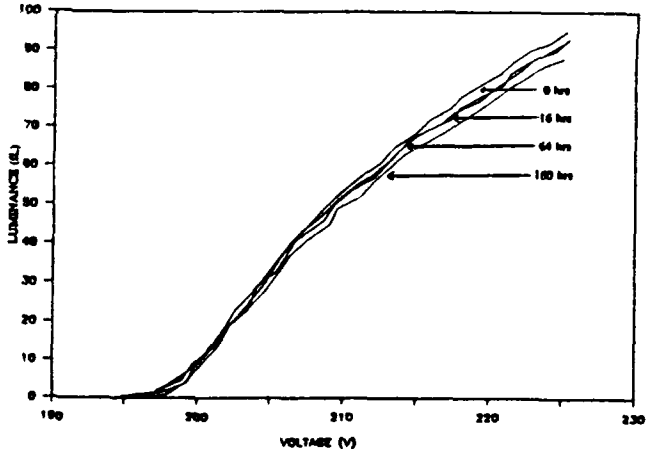


Figure 2. B-V curves for sulfur exposed sample - aging time extrapolated to 60 Hz.

The second test involves matrix panels with standard commercial monitors. This method of testing allows evaluation of the structure in an actual commercial application. This method of evaluation, by using actual monitors, can establish the effectiveness of the processing changes when panels are used in a typical application. In this test, portions of a full-sized panel are operated continuously for an extended period of time (1000 hours for the work reported herein). After various aging intervals, the panel is momentarily turned on and the brightness of the aged and unaged portions of the panels are measured. The differential brightness due to aging (B_{higher}/B_{lower}-1) is calculated for intermediate brightnesses after multiple aging intervals up to 1000 hours. For comparative purposes, the

ISSN0097-0966X/91/0000-074-$1.00 + .00 © 1991 SID

brightness difference at the voltage with the worst differential brightness (low brightness levels of about 20% of maximum brightness) are plotted as a function of aging time. This type of data reflects the performance of the display in a gray scale application. Such plots for a traditional and a sulfur exposed structure are shown in Figure 3.

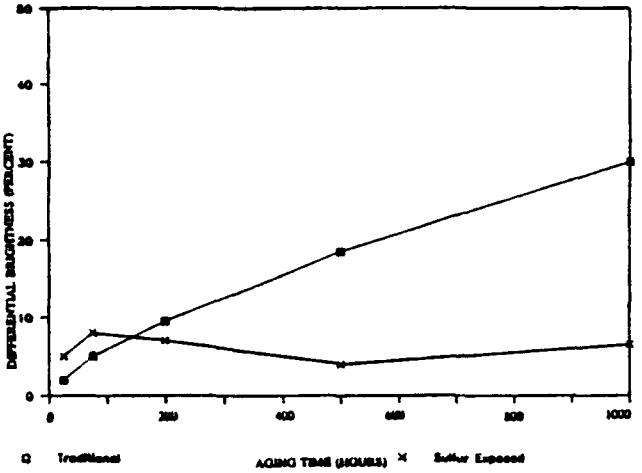


Figure 3. Differential brightness in monitors.

Simultaneous with these gray scale measurements, aging at the full "on" and "off" points, which we denote as the latent image, is measured for performance evaluation in "on"-"off" applications. The traditional structure shows a maximum of 5% aging at the "on" brightness level and below 0.1 fL at the "off" level. These changes are at the threshold of our measurement capability; therefore, improvements with the sulfur exposure could not be accurately measured.

Discussion

Proposed Aging Mechanism

The observed ACTFEL instabilities are most likely associated with changes in the space charge distribution. The external AC field perturbs the internal charge distribution, and hence the internal field, as a function of aging time. We propose an aging mechanism in which positive charges accumulate at the interface. This positive charge is balanced by an equal amount of negative charge which we presume resides in the ZnS bulk. According to this explanation, the unaged device is described by an equilibrium energy band diagram as shown in Figure 4a, while the energy band diagram of the aged device is modified as indicated in Figure 4b. If it is assumed that all the energetic electrons for luminescence originate at the ZnS/insulator interface, it is clear that the charge distribution illustrated in Figure 4b results in an increase in the threshold voltage as observed experimentally.

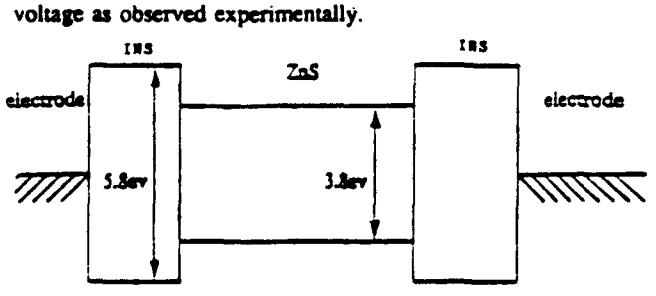


Figure 4a. Energy band diagram for unaged sample.

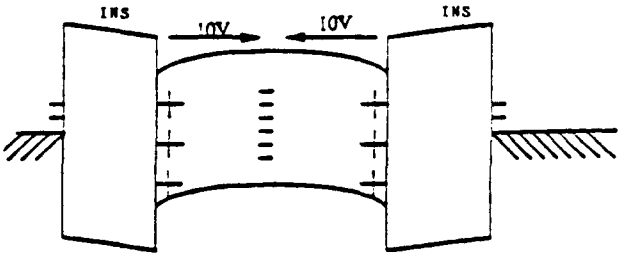


Figure 4b. Energy band diagram for aged sample.

The space charge responsible for the observed ACTFEL instability is attributed to the migration of sulfur vacancies in the ZnS layer and accumulation at the ZnS / insulator interface. Sulfur vacancies act as double donors. When ionized, the vacancy is positively charged[8]. The sulfur vacancy mechanism is consistent with the charge distribution illustrated in Figure 4b, if the sulfur vacancy concentration increases near the interfaces with aging.

Instability Mechanism

The proposed electrostatic ACTFEL instability model accounts for Vth increases with aging time and for the fact that the B-V curve shifts in a non-rigid fashion (i.e., the brightness at higher voltages stays about constant with aging while Vth increases, as shown in Figure 1).

Vth increases with aging time because of an increased tunneling barrier due to positive charge at the interface, as shown in Figure 4b. Therefore, a higher external voltage is required to reach the threshold field. It is possible to calculate the amount of transferred interface charge required to obtain a 10 volt increase in Vth. It can be shown that this ten volt shift can be caused by as little as 0.01% sulfur vacancies distributed over a distance of 100Å near the ZnS interface[9]. Such a low level of intrinsic defects precludes the use of analytical measurements to corroborate our instability model. Typical analytical methods such as Auger Electron Spectroscopy (AES), Rutherford backscattering (RBS) and secondary Ion Mass Spectroscopy (SIMS) were employed in our analysis but did not show any evidence for a change in the stoichiometry which could prove or disprove our model. These tests were performed on untreated as well as treated samples before and after aging. It appears that electrical or optical measurements are best suited to the assessment of ACTFEL instabilities.

In a previous publication[5], improvements in the stability of ZnS:Mn ACTFEL devices using CaS buffer layers were reported. It was proposed that the improvement was due to the ability of CaS to remove sulfur vacancies from the ZnS layer rather than CaS acting as a buffer against ion movement from the insulators. This postulate was supported by observations that the location of CaS on only one side of the ZnS layer did not change the stability compared with having CaS on both sides of the ZnS layer.

In follow-up work reported here, we attempted to test this postulate by excluding the buffer layer and increasing the sulfur content of the ZnS film by heat treatment in sulfur or H_2S. As shown in Figures 2 and 3, this sulfur treatment resulted in significant improvements in the device stability. Furthermore, these sulfur treatments showed effects very similar to the CaS buffer layer in increasing the device Vth and increasing the

brightness in addition to improving the stability. These sulfur exposure results are further evidence that sulfur vacancies are intimately related to the ACTFEL device stability.

Conclusion

Improved device stability is demonstrated through sulfur exposure of test structures as well as commercial devices. An electrostatic and atomistic mechanism for the ACTFEL instability in traditional device structures is proposed. The charged S-vacancies can change the internal device electrostatics through their accumulation at the interfaces during device operation. This work shows the viability of utilizing an ACTFEL display in an application with voltage-controlled intensity.

References

1. Y. Kanemori, M. Katayama, K. Nakazawa, H. Kato, K. Yano, Y. Fukuoka, Y. Kanatani, Y. Ito, M. Hijikigawa, SID 90 Digest, 408 (1990).

2. S. Faria, J. Mayo, SID 90 Digest, 246 (1990).

3. R.T. Flegal, C.N. King, SID 86 Digest, 177 (1986).

4. M. Nishikawa, T. Matsuoka, T. Tohda, Y. Fujita, J. Kuwata, A. Abe, SID 88 Digest, 19 (1988).

5. I. Khormaei, J.F. Wager, C.N. King, SID 89 Digest, 65 (1989).

6. A. Vecht, and S. Chadha, SID82 Digest, 16 (1982).

7. K. Tanaguichi, K. Tanaka, T. Ogura, Y. Kakihara, S. Nakajima, and T. Inoguchi, Proceedings of the SID, 26 (1985).

8. F.A. Kröger, The Chemistry of Imperfect Crystals, Vol. II, North-Holland (1974).

9. I. Khormaei, C.N. King, R.E. Coovert and J.F. Wager, Proceedings of the SID, to be published (1991).

Acknowledgement

The authors would like to thank Dr. Dave Morton and Mr. James Koh of Army LABCOM for analytical measurements. Assistance from Dr. Sey-Shing Sun in preparing the H2S annealed devices is also appreciated. This work was partially funded through U.S. Army Contract No. DAAL01-88-C-0857. John F. Wager wishes to acknowledge the Air Force Office of Scientific Research under Contract No. AFOSR 89-0309 for support of this work.

Electrical Characterization and Modeling of Alternating-Current Thin-Film Electroluminescent Devices

James D. Davidson, *Student Member, IEEE*, John F. Wager, *Member, IEEE*, Ron I. Khormaei, Christopher N. King, and Richard Williams

***Abstract*—Electrical characterization of evaporated ZnS:Mn alternating-current thin-film electroluminescent (ACTFEL) devices is accomplished by capacitance-voltage (C-V) analysis. Interpretation of these C-V characteristics is aided by SPICE modeling and by electrical characterization of an ideal ACTFEL device constructed from discrete components, based on a simple equivalent circuit for the ACTFEL device. Various features of the C-V curve are ascribed to equivalent circuit parameters and associated device physics parameters.**

I. Introduction

ELECTRICAL characterization of alternating-current thin-film electroluminescent (ACTFEL) devices is usually accomplished via charge-voltage (Q-V) [1]-[7] or current-voltage (I-V) [1], [5], [8], [9] analysis. We have recently described an alternative approach, the capacitance-voltage (C-V) technique [10] for analysis of the electrical properties of ACTFEL devices. The primary purpose of this paper is to compare C-V characteristics of ZnS:Mn ACTFEL devices to those obtained from SPICE simulation. The second purpose of this paper is to compare experimental and SPICE-simulated C-V characteristics to C-V curves obtained from an ideal ACTFEL device which is constructed from discrete components based on a simple equivalent circuit model of the ACTFEL device.

II. Experimental Approach

A. ACTFEL Device Structure

The ACTFEL devices have an evaporated ZnS:Mn active phosphor layer which is sandwiched between two sputtered silicon oxynitride insulator layers. Aluminum and indium-tin oxide (ITO) electrodes are employed as contacts.

Manuscript received March 11, 1991; revised October 7, 1991. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR 89-0309. The review of this paper was arranged by Associate Editor W. F. Kosonocky.

J. D. Davidson and J. F. Wager are with the Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, OR 97331.

R. I. Khormaei and C. N. King are with Planar Systems, Inc., Beaverton, OR 97006.

R. Williams is with the David Sarnoff Research Center, Princeton, NJ 08543-5300.

IEEE Log Number 9106906.

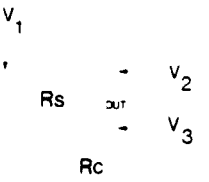


Fig. 1. Circuit used for ACTFEL electrical characterization.

B. Electrical Characterization

Electrical analysis is accomplished using the circuit shown in Fig. 1. An ac waveform generator (Wavetek model 275) with a small duty cycle is used to drive a series resistor R_s, the ACTFEL device, and a current sense resistor R_c. R_s is chosen to be large compared to R_c (typically $R_s \approx 1.5$ kΩ and $R_c \approx 10$ Ω) so that the voltage drop across R_c is negligible. The three voltages indicated in Fig. 1; v_1, v_2, and v_3; correspond to the driver, the ACTFEL device, and the current sense voltages, respectively.

The standard waveform employed in this work is indicated in Fig. 2. The waveform is symmetric with bipolar pulses of trapezoidal shape with 5-μs rise and fall times and a pulsewidth of 30 μs where the pulsewidth is defined as the duration in which the pulse is at its maximum amplitude. The frequency of the waveform is 1 kHz.

Voltages $v_1(t)$, $v_2(t)$, or $v_3(t)$ are obtained by sampling using a Tektronix model 7854 digitizing oscilloscope. An oscilloscope time base of 2 μs/division and 128 points/waveform gave best results. For the work described in this paper, only $v_2(t)$ and $v_3(t)$ are normally measured.

Using the circuit shown in Fig. 1, $i(t)$, $q(t)$, $C(v_2 - v_3)$, Q-V, and C-V may be obtained as follows. The current is obtained from the voltage across the sense resistor

$$i(t) = \frac{v_3(t)}{R_c}. \tag{1}$$

The charge is evaluated as the integral of current

$$q(t) = \int_{\tau=0}^{t} i(\tau)\, d\tau. \tag{2}$$

0018-9383/92$03.00 © 1992 IEEE

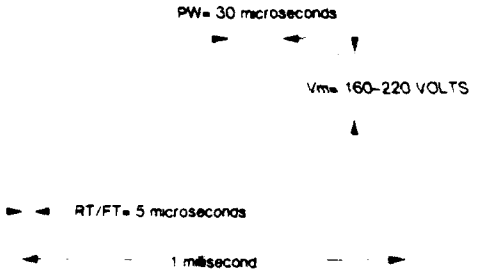


Fig. 2. Standard driver waveform used for ACTFEL electrical characterization.

The capacitance is equal to the current divided by the derivative of the voltage across the ACTFEL device [10]

$$C(v_2 - v_3) = \frac{i(t)}{d[v_2(t) - v_3(t)]/dt}. \quad (3)$$

Note that the capacitance specified by (3) is the normal differential capacitance used in conventional C-V analysis [11], [12]. Q-V and C-V are then obtained by plotting $q(t)$ and $C(v_2 - v_3)$ versus $[v_2(t) - v_3(t)]$.

Although complete electrical characterization of the ACTFEL device may be accomplished using the Sawyer-Tower [2] or series resistor [10] methods, as well as the sense resistor method of Fig. 1, we prefer the sense resistor method for the following reasons. First, the sense resistor method is not as sensitive to calibration variations as is the series resistor method where the voltage amplifiers which measure v_1 and v_2 must be very closely matched. Second, the sense resistor method requires one numerical differentiation step when the data are acquired digitally whereas two numerical differentiations are required with the Sawyer-Tower method.

C. Equivalent Circuit Analysis

A simple equivalent circuit for the ACTFEL device [2] is shown in Fig. 3. C_{i1} and C_{i2} are the insulator capacitances whereas C_p is the phosphor capacitance. The phosphor capacitance is shunted by two back-to-back Zener diodes which are used to account for conduction and field clamping in the phosphor region.

The equivalent circuit presented in Fig. 3 is quite simple and we have found that it adequately describes the main features of actual ACTEL operation. We have reached this conclusion in two ways. First, we have built an ideal ACTFEL model from the discrete components indicated in Fig. 3 and have performed electrical characterization of it which is then compared to component values of the actual device. The actual component values used in the ideal circuit made from discrete components are listed in Table I. It should be noted that this ideal ACTFEL circuit was used to calibrate our C-V measurement. This calibration procedure revealed the presence of 0.09-nF stray capacitance associated with the coaxial cable connecting the sense resistor to the ACTFEL device which had to be accounted for prior to obtaining accurate measurements of the threshold voltage, insulator, and total capacitances which then agreed with that of the known values.

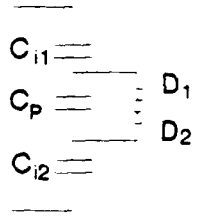


Fig. 3. Simple equivalent circuit for ACTFEL device.

TABLE I
COMPONENT VALUES FOR THE IDEAL ACTFEL MODEL BUILT FROM DISCRETE COMPONENTS

Component	Component Parameter	Component Value
C_{i1}	capacitance	2.18 nF
C_{i2}	capacitance	2.20 nF
C_p	capacitance	0.987 nF
D_1	breakdown voltage	92 V
D_2	breakdown voltage	92 V

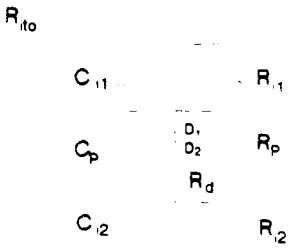


Fig. 4. Equivalent circuit for SPICE simulation of ACTFEL device.

The second approach we have taken to assess the validity of the simple equivalent circuit is to perform SPICE modeling of the ideal ACTFEL circuit subjected to various waveforms. Although SPICE simulation using the simple equivalent circuit shown in Fig. 3 reproduces the general features of the C-V curve, we find the refined equivalent circuit shown in Fig. 4 to be more accurate for SPICE modeling and also more useful for establishing parametric variation trends in C-V curves, as discussed in Section III-B 4). Five additional resistors are added to the ideal equivalent circuit of Fig. 3 to obtain the SPICE equivalent circuit of Fig. 4. R_{i1}, R_{i2}, and R_p represent the parallel resistance of the respective insulator or phosphor layers which would be associated with dc leakage through these layers. R_{ito} is a lumped resistance accounting for the nonzero series resistance of the ITO electrode. R_d we denote as the diode resistance. As we shall see, R_d is more important in establishing the width of the C-V transition than any of the other parasitic resistors. We also denote R_d as a "hot-electron resistor" since physically it is attributed to the resistance associated with hot-electron emission from interface traps. A physical interpretation of this hot-electron resistor can be obtained if it is recognized that accurate SPICE simulation of the C-V transition region requires [13] inclusion of a diode or hot-

electron capacitor C_d into the diode branch of the equivalent circuit shown in Fig. 4. $R_d C_d$ is then recognized as the time constant for hot-electron injection from interface states which gives rise to the time delay responsible for the widening of the C–V transition region. No further discussion of C_d is given herein since subtleties associated with the hot electron capacitor are beyond the scope of this paper.

III. Electrical Characterization of ACTFEL Devices

A. Q–V Analysis

A typical Q–V curve is indicated in Fig. 5. As shown in this figure, a threshold voltage, which we denote as V_{th2}, is distinguished as the intersection of the straight-line portions of the Q–V curves and indicates the external voltage at which conduction of electrons across the ZnS layer occurs. Q^e_{pol} is the externally measured polarization charge which corresponds to the charge left on the insulator after the applied waveform goes to zero and which gives rise to the internal polarization field $\mathcal{E}_{pol}$. The polarization field in the phosphor at the end of the applied waveform duty cycle is given by [6]

$$\mathcal{E}_{pol} = \frac{Q^e_{pol} C_p}{\epsilon_p A C_i} \tag{4}$$

where ϵ_p is the phosphor dielectric constant, A is the device area, and C_i is the total insulator capacitance (i.e., series combination). For ACTFEL devices in which field clamping occurs in the phosphor, Q_{cond} is the total charge transported across the phosphor while it is conducting and is the charge that gives rise to light emission.

B. C–V Analysis

A typical C–V curve is indicated in Fig. 6. C_t is the total capacitance of the ACTFEL device (i.e., a series combination of the phosphor and insulator capacitances). C_i is the insulator capacitance which is measured at larger applied voltages when the phosphor is conducting, effectively shunting the phosphor capacitance. In the following subsections various features of the C–V curve; namely, the threshold voltage, the interface trap density, the capacitance rise near the maximum applied voltage, and the parasitic resistances; are discussed in detail.

1) Threshold Voltage: It is evident from Fig. 6 that the threshold voltage is not uniquely defined. We define three threshold voltages corresponding to the onset, midpoint, and saturation of the C–V transition which we denote V_{th1}, V_{th2}, and V_{th3}, respectively. V_{th1} corresponds to the onset of emission of electrons from the most shallow filled interface traps. V_{th3} corresponds to the initiation of field clamping; that is, the external voltage at which the interface trap density is of sufficient magnitude that the phosphor field remains constant. SPICE modeling confirms these identifications of V_{th1} and V_{th3}. V_{th2} corresponds most closely to the threshold voltage found from Q–V measurements.

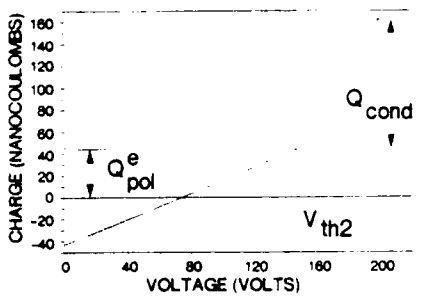


Fig. 5. Q–V curve for ZnS : Mn ACTFEL device.

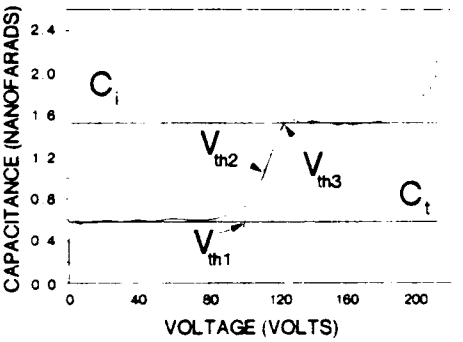


Fig. 6. C–V curve for ZnS : Mn ACTFEL device.

2) Interface State Density: The slope of the C–V curve is related to the density of interface traps in the sub-field-clamping regime as follows. Assume that the C–V curve is approximated in a piecewise-linear manner as shown in Fig. 7. In particular, assume that the C–V transition is linear and use the notation shown in Fig. 7. Since the C–V transition is assumed to be linear, we can write

$$C(V) = C_1 + \frac{C_3 - C_1}{V_3 - V_1}(V - V_1) \tag{5}$$

which is only valid in the transition region. Let Q_1 and Q_3 denote the instantaneous charge measured by a sense capacitor in a Sawyer–Tower circuit at the beginning and ending of the C–V transition and which correspond to C_1, V_1 and C_3, V_3, respectively. Recognize that

$$dQ = C(V)\, dV. \tag{6}$$

Substitute (5) into (6) and integrate from V_1 to V to obtain

$$Q(V) = Q_1 + C_1(V - V_1) + \frac{1}{2}\frac{C_3 - C_1}{V_3 - V_1}(V - V_1)^2. \tag{7}$$

The first term on the right side of (7) corresponds to the instantaneous charge at the beginning of the C–V transition while the second term is just the normal displacement charge of capacitance C_1. The last term of (7) is the externally measured conduction charge, which is equal to the interface state charge measured externally. If we evaluate (7) at $V = V_3$, we can identify the last term on the right side of (7) as the externally measured interface state

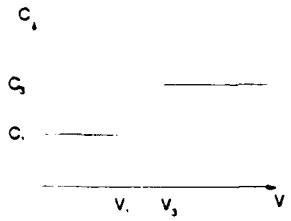


Fig. 7. Piecewise-linear C-V curve used for the interface state density derivation.

charge in the pre-clamping regime Q_{ss}^{ext} which is given by

$$Q_{ss}^{ext} = \frac{1}{2}(\Delta C)^2 \left[\frac{\Delta C}{\Delta V}\right]^{-1} \tag{8}$$

where $\Delta C = C_3 - C_1$, $\Delta V = V_3 - V_1$, and where (8) is written in a form which expresses the interface state charge in terms of the inverse of the slope of the C-V transition. If C_3 can be identified as C_i and if the right side of (8) is multiplied by $(C_i + C_p)/C_i$ to account for the difference between externally measured and internal charge [6], [14], the interface state charge in the pre-field-clamping regime Q_{ss}^{int} is (after some algebra) given by

$$Q_{ss}^{int} = \frac{C_i^2}{2}\frac{C_i}{C_p}\left[\frac{\Delta C}{\Delta V}\right]^{-1}. \tag{9}$$

Alternatively, the interface state density Q_{ss}, expressed in units of number of states per square centimeter, is as follows:

$$Q_{ss} = \frac{C_i^2}{2qA}\frac{C_i}{C_p}\left[\frac{\Delta C}{\Delta V}\right]^{-1}. \tag{10}$$

The viability of (9) and (10) is tested via SPICE simulation using the equivalent circuit of Fig. 4. This is accomplished by integrating the conduction current through R_d from the onset of conduction until field clamping, which corresponds to V_{th1} and V_{th3}, respectively. This integrated conduction charge is found to agree to within about 15% with Q_{ss}^{int}, which supports the essential validity of (9) and (10). Q_{ss} for the ACTFEL device shown in Fig. 6 is 9×10^{11} cm^{-2} using $A = 0.079$ cm^2.

Note that (10) specifies that the interface state density is inversely proportional to the slope of the C-V transition region. Thus an abrupt C-V transition implies a small density of interface states. Initially this result may seem counter-intuitive since a large interface state density is required for electron sourcing and proper ACTFEL device operation. It is important to recognize, however, that Q_{ss} corresponds to the density of interface states in the *pre-field-clamping regime*. Therefore, an interface state density with a delta function distribution would have a very abrupt C-V transition even if the strength of the delta function, corresponding to the interface state density responsible for field clamping, is exceedingly large. In contrast, a linearly increasing interface state distribution would exhibit a much more gradual C-V transition.

3) Capacitance Rise Near V_m: For a variety of reasons we believe the presence of a capacitance rise near the maximum voltage to be a feature of the measurement rather than evidence of leakage in or charge injection into the insulating layers [13]. Such an increase could be interpreted as arising from the injection of charge into the insulator such that the effective thickness of the insulator is reduced and the corresponding capacitance of the insulator is increased. Alternatively, it could be argued that the measured capacitance rise is actually due to insulator leakage which is large enough that the approximate circuit model for the insulator is a resistor in parallel with the insulator capacitance such that the apparent rise in capacitance is due to an RC effect. These two explanations can be eliminated because the distance of charge injection into the insulator and the extent of insulator leakage required to explain the capacitance rise are unrealistically large. These mechanisms can also be excluded on the basis of SPICE simulation as shown in the next section. Finally, note that the C-V curve shown in Fig. 8 which is obtained from the ideal ACTFEL device constructed from discrete components as listed in Table I also displays a rise in the capacitance near V_m, further evidence that charge injection or insulator leakage is not the cause of this capacitance increase.

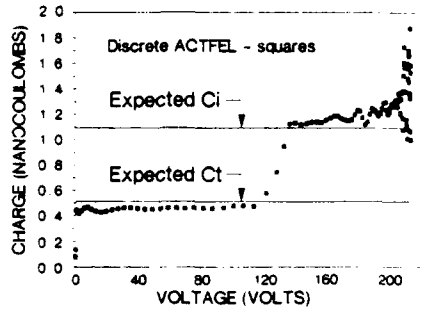


Fig. 8. Measured C-V curve of ideal ACTFEL device constructed from discrete components.

The nature of the capacitance rise can be understood by reference to Fig. 9. Recall from (3) that the capacitance is obtained from the measured current divided by the derivative of the voltage across the ACTFEL device. As indicated in Fig. 9, the capacitance rise occurs when the measured current and the voltage derivative are very close to zero. The rise occurs because the current waveform is always observed to decay more slowly than the voltage derivative.

In summary, the existence of the capacitance rise near V_m is a measurement feature associated with the fact that the measured current decays more slowly than the derivative of the voltage across the ACTFEL device.

4) Parasitic Resistances: The equivalent circuit shown in Fig. 4 and used for SPICE simulation employs five resistors in addition to the ideal equivalent circuit model of Fig. 3. A comparison between the SPICE-simulated C-V curves of an ideal device (i.e., $R_d = R_{ito} = 0$, $R_{i1} = R_{i2} = R_p = 10^{11}\ \Omega$) and a device with nominal values of circuit parameters (see Tables I and II for nominal component values used in SPICE simulations) is given in Fig. 10. Note the abruptness of the C-V transition in the sim-

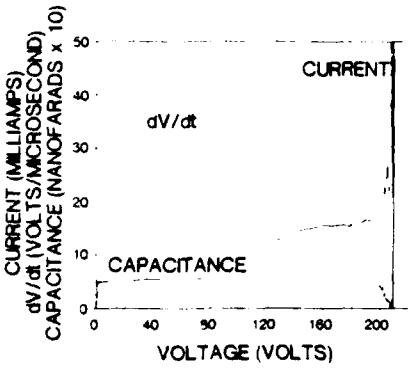


Fig. 9. Current $i(t)$ and voltage derivative $d[v_2(t) - v_1(t)]/dt$ versus time plots which help to clarify the capacitance rise near V_m.

TABLE II
NOMINAL PARASITIC RESISTANCE VALUES USED FOR SPICE MODELING

Component	Component Parameter	Component Value
R_{ito}	ITO series resistance	20 Ω
R_{i1}, R_{i2}	insulator parallel resistance	10^{11} Ω
R_p	phosphor parallel resistance	2×10^6 Ω
R_d	diode resistance	30 Ω

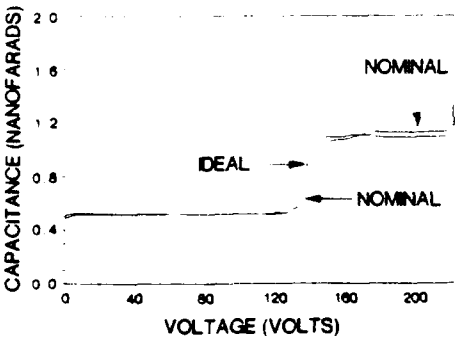


Fig. 10. SPICE simulated C-V curves of ideal ACTFEL device and a device with typical values of circuit parameters.

ulated ideal device. Also note the existence of a capacitance rise near V_m, further evidence that this C-V structure is a measurement feature and not evidence for leakage or charge injection. In contrast, the simulated C-V curve for the typical device displays a nonabrupt C-V transition, a rise in capacitance near zero voltage, and a capacitance step in the insulator capacitance regime.

To explore these parasitic resistance trends, C-V curves with parametric variations of R_{ito}, R_{i1}, R_p, and R_d around typical parameter values are shown in Figs. 11–14. Several trends emerge from an anaylsis of these figures. First, it is clear that the capacitance rise near zero voltage is associated with R_{ito} and that this rise is essentially an RC effect. It is evident from Fig. 11 that R_{ito} influences the abruptness of the C-V transition. R_{ito} is also the most important parameter in determining the capacitance step in the C_i regime of the C-V curve.

From an analysis of Figs. 12 and 13 we conclude that a decrease in the insulator or phosphor shunt resistances leads to a rigid shift in the C-V transition to higher thresh-

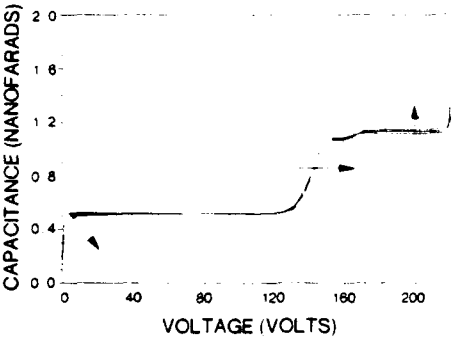


Fig. 11. C-V curves for parametric variations of R_{ito}. R_{ito} = 0, 20, 40, and 80 Ω. Arrows indicate an increasing value for R_{ito}.

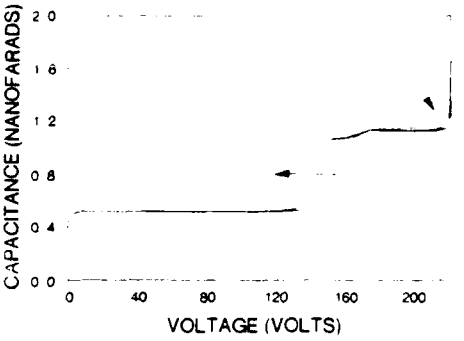


Fig. 12. C-V curves for parametric variations of R_{i1}. R_{i1} = 100 kΩ, 500 kΩ, 1 MΩ, and 100 × 10^9 Ω. Arrows indicate an increasing value for R_{i1}.

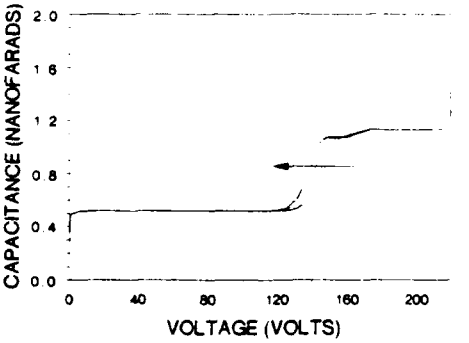


Fig. 13. C-V curves for parametric variations of R_p. R_p = 800 kΩ, 1 MΩ, 2 MΩ, 5 MΩ, and 10 MΩ. Arrow indicates an increasing value for R_p.

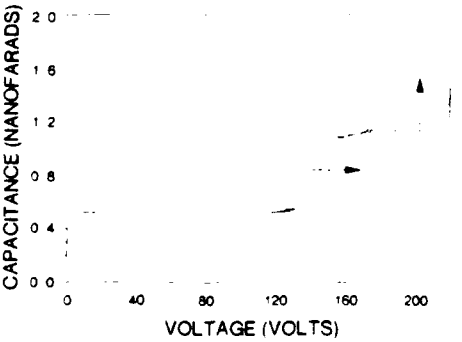


Fig. 14. C-V curves for parametric variations of R_i. R_i = 0, 30, 100, and 200 Ω. Arrows indicate an increasing value for R_i.

old voltages. A physical interpretation of this trend is that a reduction in the resistance of the phosphor or insulator layers causes leakage of the polarization charge which gives rise to a concomitant increase in the threshold voltage. We believe that the high quality of the insulators used in the ACTFEL devices tested in the present work is such that R_{i1} and R_{i2} may, for all practical purposes, be considered to be infinite. In contrast, our nominal value of R_p = 2 MΩ is not infinite and corresponds to a ZnS resistivity of $\sim 10^{12}\ \Omega \cdot$ cm; thus the phosphor shunt resistance cannot be ignored since it contributes to the establishment of the threshold voltage. Note also that the capacitance rise near V_m is affected by a small insulator shunt resistance, as shown in Fig. 12.

As indicated in Fig. 14, R_d affects the abruptness of the C–V threshold with smaller R_d yielding a more abrupt transition. Physically, R_d corresponds to a resistance associated with hot-electron injection from interface traps. Thus R_d is related to the characteristic time for electron emission from interface traps such that we interpret a smearing out of the C–V transition as arising from a higher density of interface traps with characteristic emission times of comparable magnitude to that of the driver waveform (i.e., microseconds).

Although the abruptness of the C–V transition is mainly established by the magnitude of R_d, the abruptness is also affected, but to a much lesser extent, by R_{ito}. When the abruptness of the C–V transition is determined by R_{ito} exclusively, no information about the interface state density in the sub-field-clamping regime can be obtained from the slope of the C–V curve since the width of the C–V transition is then established by a parasitic-dominated RC time constant of the ACTFEL device and not by the intrinsic, interface-trap-related RC time constant. Since the relative magnitude of R_{ito} can be readily monitored by the capacitance rise in the C–V curve near zero voltage, improper interpretation of the C–V transition can be easily avoided.

Increasing R_d also leads to a larger C_i step in the insulator capacitance regime of the C–V curve. Thus R_{ito} and R_d are the two most important parameters in determining the C_i step in the insulator capacitance regime. It should be noted that we often, but not always, observe such a C_i step in real C–V curves. It is evident from Figs. 11 and 14 that the C_i step depends mainly on R_{ito} and R_d.

This C_i step is the result of using an inadequate model for the ACTFEL device when R_{ito} and R_d are nonzero. This can be understood by referring to (3) and recognizing that this defining equation of the C–V technique implicitly assumes that the ACTFEL device is adequately modeled as a two-terminal parametric capacitor. If R_{ito} and R_d are nonzero, this implicit assumption is no longer correct. For example, if R_{ito} is nonzero, a more accurate model for the ACTFEL device would be a series circuit consisting of R_{ito} and a parametric capacitor specified [13] by an equation analogous to (3).

In our view, the C–V technique is still of much diagnostic utility, even when the most appropriate parametric device model is not employed, because the presence of a C_i step alerts one to the presence of a nonzero parasitic resistance in the ACTFEL device under test. Care must be taken, however, in evaluating C_i from a C–V curve which possesses a C_i step. As shown in Figs. 11 and 14, the actual insulator capacitance C_i is intermediate of the two constant values of capacitance in the insulator capacitance regime which define the insulator step.

IV. Conclusions

In this paper, we have demonstrated the utility of the C–V technique for the electrical characterization of ACTFEL devices. The threshold voltage, insulator capacitance, total capacitance, and the density of interface traps in the sub-field-clamping regime may be extracted from the measured C–V curve. The capacitance rise in the C–V curve near the maximum applied voltage has been shown to be a measurement feature not due to leakage or charge injection into the insulator.

SPICE modeling in conjunction with a simple equivalent circuit model for the ACTFEL device has been employed in order to assist in the interpretation of C–V curves. The following trends arise from SPICE modeling:

1) The capacitance rise near zero voltage is due to a nonzero value of R_{ito}, the parasitic series sheet resistance of the ITO conducting layer. This is a simple RC effect.

2) The abruptness of the C–V transition is most sensitive to R_d but also depends on R_{ito}. Physically, R_d is related to the characteristic time for the emission of hot electrons from interface traps. A nonabrupt C–V transition caused by a nonzero value of R_d is related to the density of interface states in the sub-field-clamping regime. In contrast, a nonzero value of R_{ito} can also give rise to some lack of abruptness in the C–V transition but this is associated with the overall RC time constant of the ACTFEL device and is not due to the interface traps.

3) The C–V transition shifts rigidly to larger threshold voltage with smaller R_p, R_{i1}, or R_{i2}. This shift is due to polarization charge leakage and a concomitant reduction in the polarization field in the phosphor.

4) A second, small capacitance rise in the insulator capacitance regime of the C–V curve depends strongly on R_{ito} and R_d but weakly on R_p, R_{i1}, and R_{i2}. This capacitance rise is a consequence of modeling the ACTFEL device as a simple parametric capacitor without adequately accounting for the internal resistance of the device under test.

Acknowledgment

The authors wish to thank E. Bringuier for his derivation of the interface state density which is included in this paper and for many constructive suggestions. The authors are indebted to M. Åberg for pointing out that the capacitance rise in the insulator region of the C–V curve is an RC effect.

References

[1] Y. S. Chen and D. C. Krupka, "Limitation imposed by field clamping on the efficiency of high-field ac electroluminescence in thin films," *J. Appl. Phys.*, vol. 43, pp. 4089–4096, 1972.

[2] D. H. Smith, "Modeling a.c. thin-film electroluminescent devices," *J. Lumin.*, vol. 23, pp. 209–235, 1981.

[3] K. W. Yang, "Electrical and optical measurements on a.c. thin-film electroluminescent devices," Ph.D. dissertation, Oregon State University, Corvallis, 1981.
[4] P. M. Alt, "Thin-film electroluminescent displays: Device characteristics and performance," in *Proc. SID*, vol. 25, 1984, pp. 123-146.
[5] Y. A. Ono, H. Kawakami, M. Fuyama, and K. Onisawa, "Transferred charge in the active layer and EL device characteristics of TFEL cells," *Japan. J. Appl. Phys.*, vol. 26, pp. 1482-1492, 1987.
[6] E. Bringuier, "Charge transfer in ZnS-type electroluminescence," *J. Appl. Phys.*, vol. 66, pp. 1314-1325, 1989.
[7] L. E. Tannas, *Flat-Panel Displays and CRTs*. New York: Van Nostrand Reinhold, 1985, ch. 8.
[8] G. O. Müller, R. Mach, B. Selle, and G. Schulz, "Measuring on thin film electroluminescent devices," *Phys. Status Solidi (a)*, vol. 110, pp. 657-669, 1988.
[9] K. Neyts and P. DeVisschere, "Measuring the current-voltage characteristics of thin-film electroluminescent devices," *Acta Polytech. Scandinav.*, no. 170, pp. 291-294, 1990.
[10] R. C. McArthur, J. D. Davidson, J. F. Wager, I. Khormaei, and C. N. King, "Capacitance-voltage characteristics of alternating-current thin-film electroluminescent devices," *Appl. Phys. Lett.*, vol. 56, pp. 1889-1891, 1990.
[11] F. Kouril and K. Urba, *Nonlinear and Parametric Circuits*. Chichester, UK: Ellis Horwood, 1988.
[12] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. New York: Wiley, 1981.
[13] J. D. Davidson, "Capacitance-voltage analysis, SPICE modelling, and aging studies of AC thin-film electroluminescent devices," M.S. thesis, Oregon State University, Corvallis, 1991.
[14] R. Mach and G. O. Müller, "Efficiency and saturation in AC thin film EL structures," *Phys. Status Solidi (a)*, vol. 81, p. 609, 1984.

James D. Davidson (S'87) received the B.S. degree in electrical engineering from Oregon State University, Corvallis, in 1989. He is currently working towards the M.S. degree in electrical engineering at Oregon State University.

John F. Wager (S'80-M'81) received the B.S. degree in engineering physics in 1977 from Oregon State University, Corvallis, and the M.S. and Ph.D. degrees in electrical engineering from Colorado State University, Fort Collins, in 1977 and 1981, respectively.

From 1982 to 1984, he was with the Chemical Physics Department at Hughes Research Laboratories. In 1984, he joined the Department of Electrical and Computer Engineering at Oregon State University where he is presently an Associate Professor. His current research interests include alternating-current thin-film electroluminescence and defects in compound semiconductors.

Ron I. Khormaei received the B.S.E.E. and M.S.E.E. degrees from Oregon State University, Corvallis, in 1988 and 1989, respectively, where his work concentrated on electroluminescent device physics.

Currently, he is a Research Scientist at Planar Systems, Inc., Beaverton, OR, where he is continuing work on characterization of ACTFEL devices and development of color displays.

Christopher N. King received the B.A. degree in physics from the University of California, Davis, 1966, the Ph.D. degree in applied physics from Stanford University in 1972.

Some of his present responsibilities include leadership of technology thrust for research, development and manufacturing. Program activities include EL materials and processing, EL addressing, chip-on-glass packaging technology, automated assembly, and high-volume panel manufacturing technology. He is a member of the Corporate Strategic planning group at Planar Systems, Inc., Beaverton, OR, for setting goals and managing present operations and future growth. His prior experience includes work at Tektronix as department manager, solid state research laboratory. He is the author or co-author of over 30 technical publications and presentations as well as four U.S. patents.

Richard Williams received the B.A. degree in chemistry from Miami University, Oxford, OH, in 1950 and the Ph.D. degree in physical chemistry from Harvard University, Cambridge, MA, in 1954.

He has been a technical staff member at the David Sarnoff Research Center, Princeton, NJ, since 1958. His research interests include insulator physics, liquid crystals, and physical properties of surfaces.

Aging studies of evaporated ZnS:Mn alternating-current thin-film electroluminescent devices

J. D. Davidson,[a] J. F. Wager, and S. Kobayashi[b]
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331-3211

(Received 16 October 1991; accepted for publication 5 January 1992)

A study of the aging characteristics of evaporated ZnS:Mn alternating-current thin-film electroluminescent (ACTFEL) devices is undertaken by monitoring the capacitance-voltage (C-V) characteristics at various temperatures as a function of aging time. Short-term ACTFEL aging is characterized by a rigid shift in the C-V curve to higher turn-on voltage with aging time. Additionally, the insulator and phosphor capacitances are found to be independent of aging time, the internal phosphor threshold voltage increases slightly with aging time, and the conduction and polarization charges are observed to decrease with aging time. The activation energy for ACTFEL aging is estimated to be about 0.2 eV. These experimental observations lead to a picture for ACTFEL aging in which atomic rearrangement at insulator/phosphor interfaces gives rise to the formation of deep level, fixed charge states. Transported electrons subsequently trapped in these deep levels reduce the amount of charge available for conduction with a concomitant reduction in the polarization charge. The reduction in the polarization charge is responsible for the observed increase in the turn-on voltage with aging.

I. INTRODUCTION

The importance of the aging characteristics of alternating-current thin-film electroluminescent (ACTFEL) devices is underscored by the fact that device stability and aging were major focal points in the paper by Inoguchi *et al.*[1] which ushered in the modern age of ACTFEL technology. Much work has subsequently been reported[2–15] regarding ACTFEL aging characteristics, mechanisms, and process modifications to improve device stability. Most of this work has involved characterization of the luminescent properties of the ACTFEL devices as a function of aging time.

The purpose of the work discussed herein is to report an investigation of the aging properties of evaporated ZnS:Mn ACTFEL devices which were fabricated by Planar Systems. This study is unique in that the aging characteristics are monitored, as a function of temperature and aging time, using the capacitance-voltage (C-V) technique.[16–18] The C-V technique offers several advantages for ACTFEL aging studies. First, the technique can be readily automated. Second, a wealth of device physics data can be acquired which leads to a better understanding of the internal electrostatic modifications associated with aging and, hence, of the physical mechanisms of aging. The obvious disadvantage of the C-V technique is that it exclusively monitors the electrical properties of the ACTFEL whereas the device performance must ultimately be assessed optically.

Temperature-dependent C-V aging studies yield an aging activation energy of approximately 0.2 eV. ACTFEL aging is also characterized by a turn-on voltage which increases with aging time, a C-V transition which shifts rigidly with aging time, phosphor and insulator capacitances which are independent of aging time, a phosphor field which increases slightly with aging time, and conduction and polarization charges which decrease with increasing aging time. These experimental observations lead to a model for ACTFEL aging in which hot-electron-mediated atomic migration near the insulator/phosphor interfaces gives rise to the creation of defect complexes which act as deep level traps and increase the interface fixed charge density. The atomic rearrangement is envisaged to occur very close to the interface (within perhaps 100 Å), to most likely involve the presence of sulfur vacancies, and to probably occur by nearest-neighbor hopping.

II. EXPERIMENTAL PROCEDURE

The ACTFEL devices are fabricated at Planar Systems and consist of an evaporated ZnS:Mn active phosphor layer which is sandwiched between two sputtered silicon oxynitride insulator layers. Aluminum and indium-tin oxide (ITO) electrodes are employed as contacts.

The C-V technique has been previously described in detail.[16–18] Briefly, C-V analysis is accomplished using the circuit shown in Fig. 1. An arbitrary waveform generator (Wavetek model 275) in conjunction with a high-voltage operational amplifier (Apex PA-85) generates the small duty cycle bipolar pulse waveform which drives a series resistor, R_s, the ACTFEL device, and a current sense resistor, R_c, as shown in Fig. 1. R_s is chosen to be 1.25 kΩ and R_c is chosen to be 10 Ω. $v_2(t)$ and $v_3(t)$ are obtained using a Tektronix model 7854 digitizing oscilloscope.

The standard waveform consists of symmetric, bipolar pulses of trapezoidal shape with 5-μs rise and fall times and a pulse width of 30 μs where rise and fall times are defined as the time between 0% and 100% of the maximum am-

[a] Present address: Intel, Hillsboro, OR 97124.
[b] Permanent address: Tsukuba Research Laboratory, Nippon Sheet Glass, Tsukuba, Japan.

 0021-8979/92/084040-09$04.00

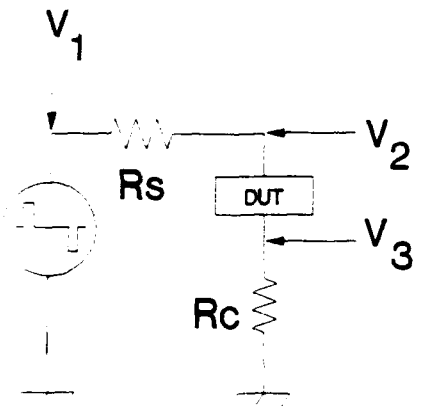


FIG. 1. Circuit used for ACTFEL electrical characterization.

plitude and the pulse width is defined as the duration during which the pulse is at its maximum amplitude. The frequency of the waveform is 1 kHz.

Referring to Fig. 1, the current through the ACTFEL device is obtained from the voltage across the sense resistor, R_c

$$i(t) = v_3(t)/R_c. \tag{1}$$

The capacitance is equal to the current divided by the derivative of the voltage across the ACTFEL device, so that

$$C(v_2 - v_3) = \frac{i(t)}{d[v_2(t) - v_3(t)]/dt}. \tag{2}$$

The C-V curve is obtained by plotting $C(v_2 - v_3)$ vs $[v_2(t) - v_3(t)]$.

A typical C-V curve is shown in Fig. 2. C_t refers to the total capacitance prior to breakdown while C_i is the insulator capacitance. Since the C-V transition is nonabrupt, we denote three turn-on voltages, $V_{to\,1}$, $V_{to\,2}$, $V_{to\,3}$, which refer to the onset of conduction, the midpoint of the C-V transition, and the field-clamping voltage, respectively. $V_{to\,2}$ is found[16-18] to correspond very well to the normal Q-V threshold.

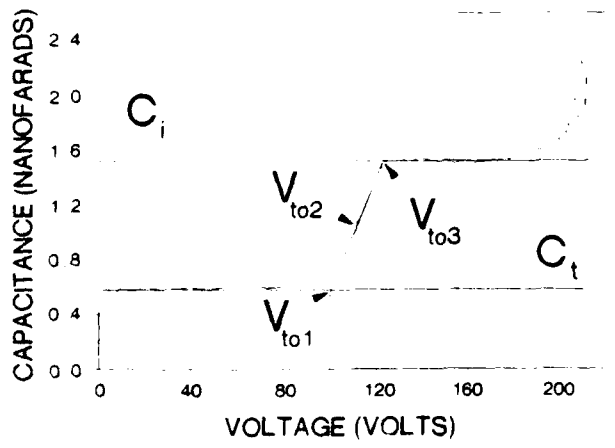


FIG. 2. C-V curve for ZnS:Mn ACTFEL device.

Note that we now choose to refer to these C-V voltages as "turn-on" voltages in contrast to our previous designation of "threshold" voltages.[16-18] We employ turn-on as a means of distinguishing these C-V voltages from the brightness-voltage (B-V) threshold voltage. B-V and C-V measurements are distinctly different as are the critical voltages measured in each experiment. A B-V curve is measured by ramping the maximum applied voltage, V_m, and monitoring the device brightness; the B-V threshold voltage corresponds to V_m at which a given level of brightness is obtained. In contrast, C-V or charge-voltage (Q-V) measurements are performed at constant V_m; the critical voltages, which we now denote as turn-on voltages, correspond to voltages at which a transition occurs in the C-V or Q-V curve. The B-V threshold is unique whereas the C-V or Q-V turn-ons depend on the magnitude of V_m because V_m determines the magnitude of the polarization charge. Finally, note that the B-V threshold also corresponds to V_m at which the Q-V curve is no longer a straight line but begins to exhibit hysteresis.

The aging experiment consists of monitoring the C-V turn-on voltage, $V_{to\,2}$, with aging time and as a function of temperature. The experiment is performed using a single glass substrate with 10 aluminum dots. Aging occurs over a temperature range of -50 to 80 °C. The aging duration is 45 h using the standard waveform with a voltage amplitude of 210 V. Temperature is controlled using a Blue-M environmental furnace. Thermocouples located in the furnace near the sample record the temperature. Turn-on voltage sampling is performed automatically with a computer-controlled system using the C-V analysis technique and the aging is carried out under continuous applied field. The turn-on voltage is obtained graphically from the C-V curves generated.

All of the dots undergo a room-temperature warmup to full operating voltage to ensure self-healing of defects that would otherwise cause burnout. The warmup procedure consists of using the standard waveform at an initial voltage amplitude of 10 V and increasing the amplitude to 210 V in 3-V steps over a period of approximately 35 min. Each dot is then subsequently further aged, at 210 V, as required, to obtain an initial turn-on of $V_{to\,2} = 125$ V, to within 2%.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. ACTFEL aging experimental results

The results of ACTFEL aging experiments are summarized in Figs. 3–5 for the following temperatures: -50 °C, -10 °C, 0 °C, 20 °C, 60 °C, and 80 °C. The family of selected C-V curves shown in Fig. 3 are for the 60 °C experiment. These C-V curves show characteristics typical of all aging experiments at various temperatures. In general, it is observed that the turn-on voltage shifts rigidly with operating time and that C_i and C_t remain constant, to within experimental error, with aging time.

As can be observed from Figs. 4 and 5, the aging characteristics can be classified into four regimes: (1) incubation period, (2) logarithmic aging, (3) saturation, and (4) long-term aging.

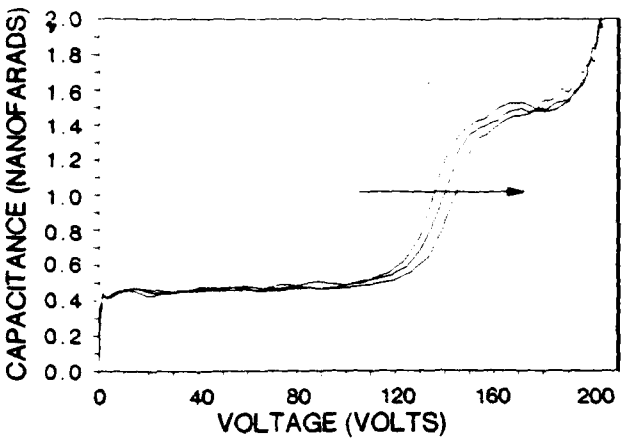


FIG. 3. *C-V* curves for 60 °C experiment. Arrow indicates direction of increasing aging time. Aging times of curves shown are 1 s, 20 min, and 9.5 h.

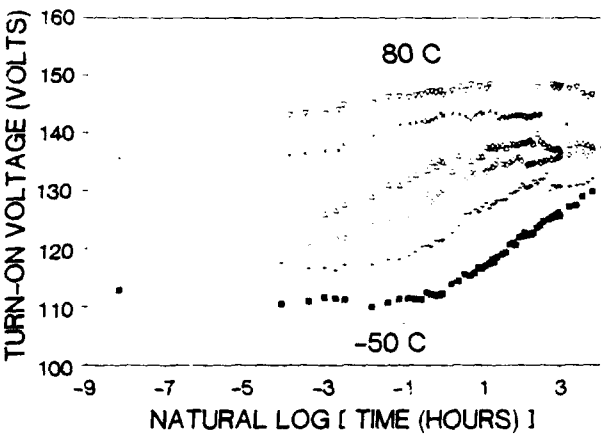


FIG. 5. Turn-on voltage, $V_{to\ 2}$, as a function of the *natural logarithm* of aging time at temperatures of -50, -10, 0, 20, 60, 80 °C.

At lower temperatures there is an incubation period, in which the turn-on voltage is essentially constant; this incubation period is most clearly evident in Fig. 5. This incubation period lasts 1 h at the lowest temperature (-50 °C) and eventually disappears at and above room temperature (20 °C). The incubation period is followed by a period in which the turn-on voltage increases logarithmically with aging time. Next, the turn-on voltage approaches a saturated value which is temperature-dependent. Finally, the last aging regime is denoted long-term aging in which the turn-on voltage decreases slowly with increasing aging time. The logarithmic and saturation aging regimes are taken to collectively comprise what we denote the short-term aging characteristics, which is the focus in the remainder of this paper. The incubation period is tentatively attributed to the fact that the warm-up and preaging occur at a higher temperature than that of the aging. Long-term aging kinetics will be the subject of future study. Thus, our present study of the aging kinetics will focus on short-term aging in which the kinetics are found, or extrapolated, to be logarithmic or saturated.

During the aging experiment it is observed that the turn-on voltage is temperature-dependent. Figure 6 shows the temperature dependence of a fully aged dot. The observed change in turn-on voltage is attributed completely to temperature since the device had been well aged and the experiment was conducted in a short period of time. Figure 6 clearly shows a linear dependence of the turn-on voltage on temperature. The initial and final turn-on voltage for each aging temperature, as shown in Fig. 7, also indicates a linear dependence of the turn-on voltage with temperature.

The cause of the temperature dependence of the turn-on voltage is uncertain. The temperature dependence of the insulator capacitance was checked from 20 to 95 °C and found to vary by only 4%. This small variation in the dielectric constant is insufficient to explain the observed turn-on voltage temperature-dependence.

B. Activation energy extraction from aging kinetics

If the incubation and long-term aging periods are temporarily ignored, ACTFEL aging experiments show an in-

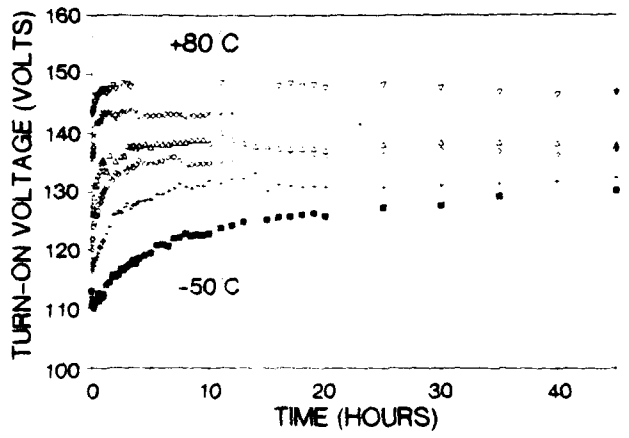


FIG. 4. Turn-on voltage, $V_{to\ 2}$, as a function of aging time at temperatures of -50, -10, 0, 20, 60, and 80 °C.

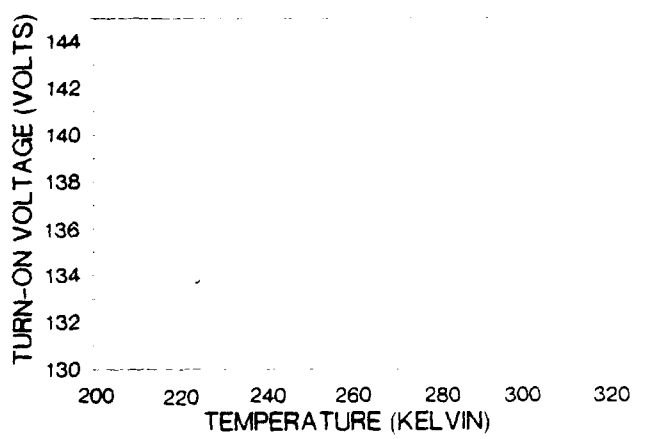


FIG. 6. Turn-on voltage, $V_{to\ 2}$, as a function of the temperature for a fully aged dot.

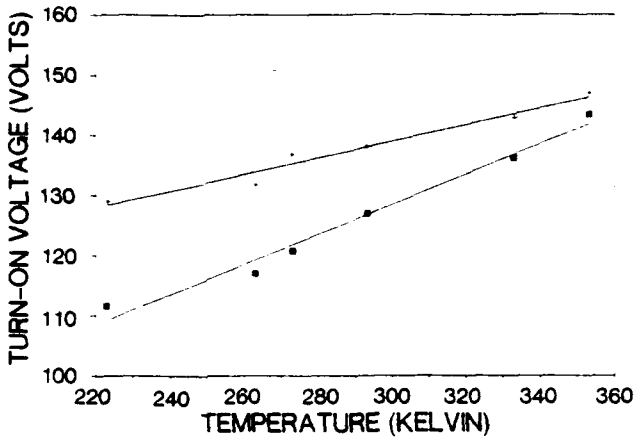


FIG. 7. Turn-on voltage, $V_{to\,2}$, as a function of temperature for initial (lower curve) and final aging times.

crease in the turn-on voltage which depends logarithmically on the aging time with a subsequent turn-on voltage saturation. Further, there is a temperature dependence of the turn-on voltage that is relatively independent of aging. The following equation is found to fit the ACTFEL aging data in the logarithmic and saturation regions and to account for the temperature dependence of the turn-on voltage:

$$\Delta V_{to}(t) = \Delta V_{to}^{SAT}(T)[1 - \exp(-\alpha t)], \tag{3}$$

where $\Delta V_{to}(t)$ is the measured change in turn-on voltage at a given temperature, $\Delta V_{th}^{SAT}(T)$ is the difference between the saturation turn-on voltage and the initial turn-on voltage at a given aging temperature, and α is the rate constant characteristic of the ACTFEL aging process. Equation (3) can be rearranged to yield

$$-\alpha t = \ln\left[1 - \frac{\Delta V_{to}(t)}{\Delta V_{to}^{SAT}(T)}\right]. \tag{4}$$

Therefore, a plot of the right-hand side of (4) versus time should result in a linear relationship in time, the slope of which corresponds to the rate constant, α. If $\ln \alpha$ is extracted for each temperature and plotted versus the inverse

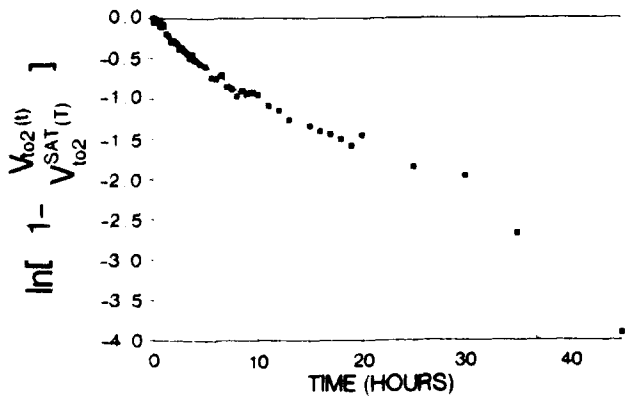


FIG. 8. Normalized shift in turn-on voltage as a function of aging time at a temperature of -50 °C.

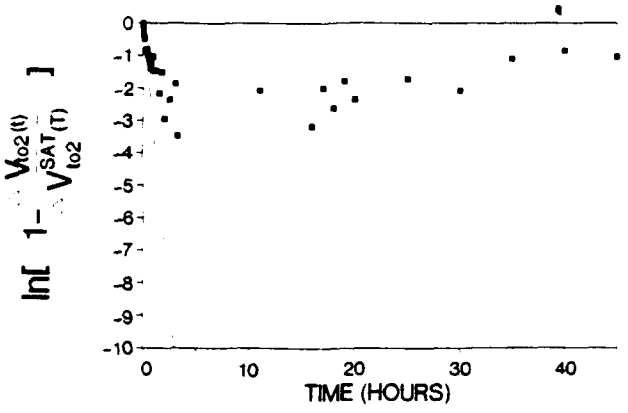


FIG. 9. Normalized shift in the turn-on voltage as a function of aging time at a temperature of 80 °C.

of the absolute temperature, the slope of such a plot yields the activation energy for ACTFEL aging.

Each of the six sets of aging data is plotted in accordance with Eq. (4). Two of these normalized turn-on voltage difference curves are given in Figs. 8 and 9. A value for $\Delta V_{to}^{SAT}(T)$ is found for each temperature from the experimental data and the rate constant, α, is determined from best fit of the linear region of these normalized plots. The natural algorithm of the rate constant for each temperature is then plotted as a function of inverse absolute temperature to obtain an Arrhenius plot as shown in Fig. 10. The slope of this line, which corresponds to the turn-on voltage activation energy for aging, has a value of approximately 0.2 eV.

IV. DISCUSSION

A. Aging electrostatics

C-V curves for the short-term, logarithmic aging regime are characterized by: (1) C_i and C_i do not vary appreciably with aging time, (2) the C-V transition region shifts rigidly with aging time, and (3) the turn-on voltage increases with aging time. Consideration (1), that C_i and C_i do not vary appreciably with aging time, implies that the centroid of the space charge giving rise to the turn-on volt-

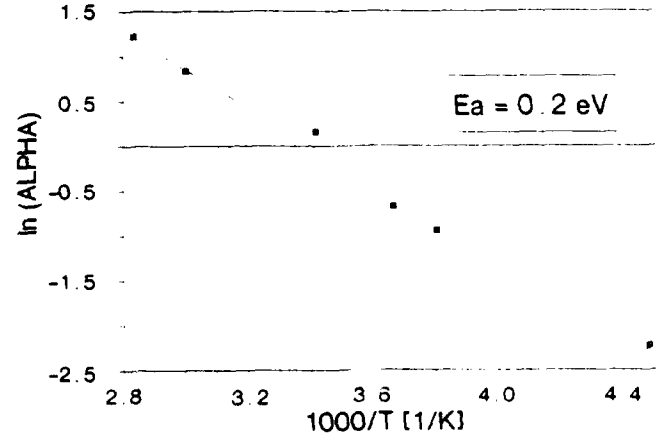


FIG. 10. Arrhenius plot of the ACTFEL aging experimental data.

age must be located near the SiON/ZnS interfaces. The maximum distance from the interface that the centroid of trapped charge is located is denoted as Δd and is given by:

$$\Delta d = \epsilon A \Delta C / C^2, \tag{5}$$

where ϵ is the dielectric constant for the insulator or phosphor, whichever is appropriate, ΔC is the capacitance measurement uncertainty, and C^2 is the insulator or phosphor capacitance, whichever is appropriate. Using $\Delta C_t = \pm 0.05$ nF and $\Delta C_i = \pm 0.1$ nF in Eq. (5) leads to $\Delta d_{ZnS} = 680$ Å and $\Delta d_{ins} = 200$ Å. In other words, the experimental fact that C_t and C_i are constant to within the specified measurement uncertainties implies that the centroid of the space charge giving rise to the threshold voltage is located within 200 Å of the interface in the insulator or within 680Å of the interface in the ZnS. Indeed, these distances are actually upper limits and it is likely that the trapped charge resides much closer than this to the interface.

Consideration (2), that the C-V transition region shifts rigidly with aging time, implies that the interface state density prior to field-clamping, N_{ss}, does not change significantly with aging. This rigid C-V shift with aging implies, in accord with metal-insulator-semiconductor (MIS) C-V studies,[19] that the observed turn-on voltage shift is consistent with an increase in the fixed charge density but not with an increase in the interface state density. Note that fixed and interface state charge are both located at the interface; fixed charge is distinguished by the fact that it is unperturbed with respect to the applied bias whereas interface states may be charged or discharged by an applied bias and interface state charge contributes to the conduction current. Fixed charged gives rise to rigid shifts in the C-V curve, whereas interface state charge leads to nonrigid C-V shifts; thus, a difference in the slope of the C-V curve is a good indicator that the interface state charge has been perturbed. Although evaporated ZnS:Mn ACTFEL devices exhibit rigid C-V shifts with aging time, this is not true of all ACTFELs. In particular, ZnS:Mn ACTFEL devices grown by atomic layer epitaxy possess C-V characteristics in which the slope decreases, and hence N_{ss} increases, with aging time.

Consideration (3), that the turn-on voltage increases with aging time, can be attributed to an increase in the internal, phosphor turn-on voltage, and/or to a reduction in the polarization charge, and hence the polarization field. To assess which of these changes is responsible for the observed increase in the turn-on voltage, plots of (a) the turn-on voltage, $V_{to\ 1}$, (b) the change in polarization and conduction charges, ΔQ^e_{pol} and ΔQ_{cond}, with aging time (note that in this paper Δ will denote a change with respect to aging time), and (c) the internal phosphor electric field at turn-on, ξ_p^{to}, as a function of aging time are shown in Fig. 11. ξ_p^{to} is determined[20] from the turn-on voltage, $V_{to\ 1}$, and Q^e_{pol} as follows:

$$\xi_p^{to} = \frac{C_i V_{to1}}{d_p(C_i + C_p)} + \frac{Q^e_{pol}}{d_p C_i}. \tag{6}$$

The superscript e in Q^e_{pol} is used to emphasize that this quantity is an externally measured quantity; the corresponding internal polarization charge is denoted Q_{pol} and these two polarization charges are related as follows:[20–22]

$$Q_{pol} = \frac{C_i + C_p}{C_i} Q^e_{pol}. \tag{7}$$

Q^e_{pol} and Q_{cond} are evaluated directly from the Q-V curve as follows.[21] Q^e_{pol} is taken to be one-half of the width of the Q-V curve evaluated at an applied voltage of zero; this corresponds to the externally measured charge which

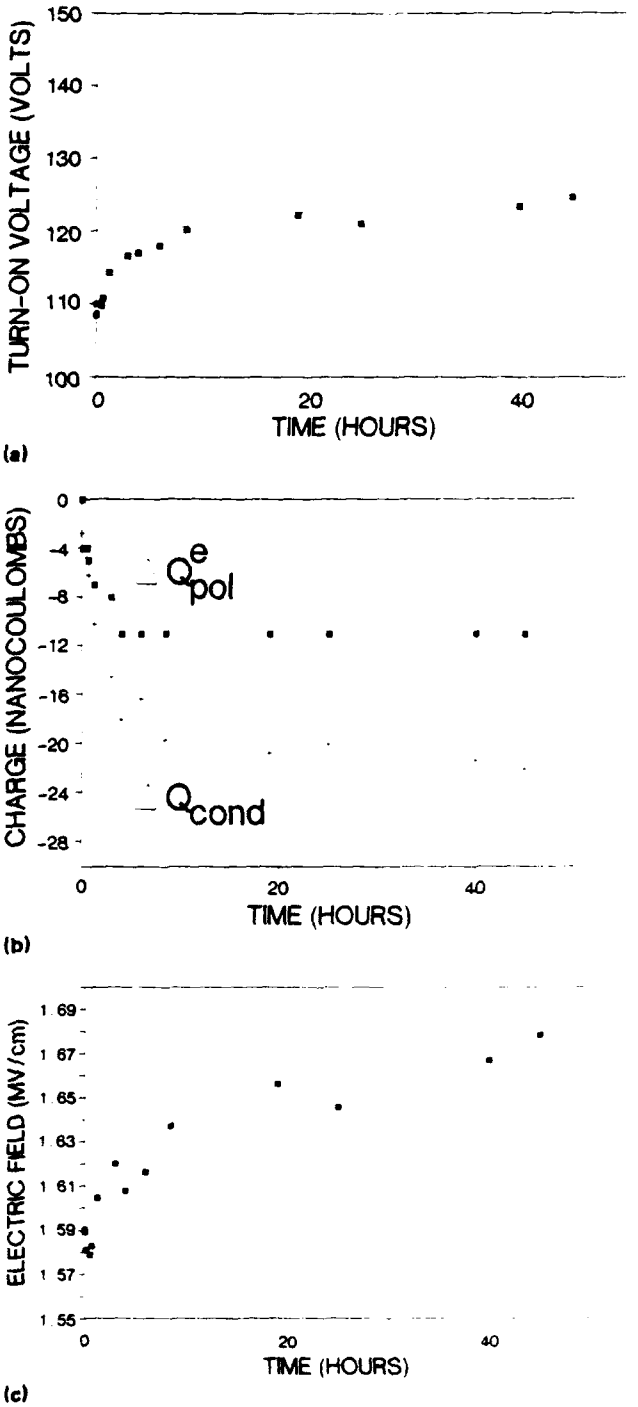


FIG. 11. Plots of (a) the turn-on voltage, $V_{to\ 1}$, (b) the change in polarization and conduction charges with aging, ΔQ^e_{pol} and ΔQ_{cond}, and (c) the internal phosphor electric field at threshold, ξ_p^{to}, as a function of aging time at a temperature of -10 °C.

is stored at the phosphor interface after the voltage pulse goes to zero. Q_{cond} is the change in charge measured from the turn-on voltage to the maximum applied voltage and corresponds to the internal charge which is transferred across the phosphor above turn-on.

It is evident from Fig. 11 that the experimentally observed increase in the turn-on voltage with aging time arises, at least partially, from a decrease in Q^e_{pol} with aging time. Note that Q_{cond} and Q^e_{pol} both decrease with aging time in the same manner; this is not unexpected[21] since if field-clamping is assumed to occur:

$$Q_{cond} = \frac{C_i + C_p}{C_i} 2Q^e_{pol}. \tag{8}$$

Also note from Fig. 11(c) that ξ^{to}_p increases slightly with respect to aging time [i.e., an increase in ξ^{to}_p of 0.09 MV/cm is evident from Fig. 11(c); the other aging temperatures yield increases in ξ^{to}_p over the duration of the aging from 0.04 to 0.11 MV/cm].

We can estimate the contribution of the change in the polarization charge, ΔQ^e_{pol}, and the change in the phosphor turn-on field, $\Delta\xi^{to}_p$, to the change in the turn-on voltage, $\Delta V_{to\ 1}$, by evaluating Eq. (6) in terms of these quantities and rearranging to yield

$$\Delta V_{to1} = \frac{C_i + C_p}{C_i}\left(d_p \Delta\xi^{to}_p - \frac{\Delta Q^e_{pol}}{C_i}\right). \tag{9}$$

Note from Fig. 11(b) that ΔQ^e_{pol} is a negative quantity such that the last term of Eq. (9) is a positive quantity which increases the turn-on voltage with aging time. The fraction of $\Delta V_{to\ 1}$ arising from the ΔQ^e_{pol} term of Eq. (9) is 0.43 ± 0.04 for the various aging temperatures. Thus, about 40% of the change in the turn-on voltage arises directly from the change in the polarization charge, and hence the polarization field, whereas approximately 60% of the change is due to the small increase in the internal, phosphor field at turn-on.

Physically, a change in the electric field, $\Delta\xi^{to}_p$, can arise from: (1) a perturbation in the tunneling barrier seen by interface states due to trapped fixed charge, (2) a shift in the density of filled interface states to deeper energies because of the trapped fixed charge (since the conduction charge which is trapped arises from the more shallow interface states, an equivalent amount of conduction charge after trapping must arise from deeper interface states), (3) a change in the interface state density with aging.

Possibility (3) does not seem as physically viable as (1) and (2). Also, the rigid shift in the C-V curve implies that at least the preclamping state density is unaffected by aging. Thus, we conclude that the observed increase in ξ^{to}_p as calculated from Eq. (6) and plotted in Fig. 11(c) arises from the trapped fixed charge due to a perturbation of the tunnel barrier and/or a shift of the filled interface states to deeper energies. Therefore, although Eq. (9) implies that $\Delta V_{to\ 1}$ can be simply resolved into distinct contributions from ΔQ^e_{pol} and $\Delta\xi^{to}_p$, physical considerations lead us to conclude that both contributions arise from the same source, namely, trapping of conduction charge into deep level, fixed charge states.

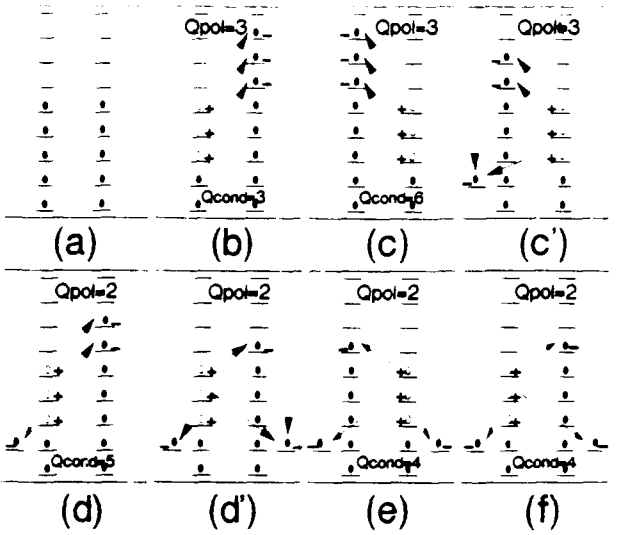


FIG. 12. Idealized electron transport and trapping sequence for an ACTFEL device with symmetric interfaces. (a) Equilibrium; charge neutrality exits, (b) negative bias; no trapping, (c) positive bias; no trapping, (c') maintain positive bias; trap 1 electron, (d) negative bias; no trapping, (d') maintain negative bias, trap 1 electron, (e) positive bias; no further trapping, (f) negative bias; no further trapping.

The fraction of conduction charge which is trapped during the aging experiment, defined as $[Q_{cond}(t=0) - Q_{cond}\ (t=45\ \text{h})]/Q_{cond}(t=0)$, is found to be between 10% and 14% for the aging temperatures investigated. Thus, trapping of a relatively small fraction of the initial conduction charge into deep level, fixed charge states gives rise to the observed increase in the turn-on voltage. Since this fraction of trapped charge is rather small, our consideration (1) argument, that the trapped charge is localized very close to the interface because the shift in the charge centroid is rather small, is weakened. In spite of the inconclusiveness of the consideration (1) argument, it is our contention that trapping of conduction charge into deep level, fixed charge states occurs very near the phosphor/insulator interface.

From the above considerations, the following picture of the aging electrostatics emerges. Some of the conduction charge originally available for transport is trapped as a function of aging time at or very close to the interfaces, giving rise to less conduction and polarization charge. Since there is less polarization charge available, the turn-on voltage increases. The electron charge is trapped at deep traps located near the interface which are not dischargeable by the applied voltage; these electron-filled deep traps contribute to fixed charge in contradistinction to interface state charge.

The details of the aging electrostatics can be better appreciated by referring to Fig. 12 which illustrates an idealized electron transport and trapping sequence for a device with symmetrical interfaces, which is consistent with our experimentally observed aging trends. Sequence (a) shows the equilibrium situation prior to the application

of a bias in which charge neutrality exists at both interfaces. The two columns of horizontal lines represent interface states at the left and right interfaces and the filled circles represent electrons occupying these interface states. Sequences (b) and (c) indicate how the interface state occupancy changes with respect to the application of positive and negative pulses if no deep trapping is presumed to occur. The empty circles represent interface states initially filled under equilibrium charge neutrality conditions but which are now not filled with electrons. + and − indicate the charge state deviation from electrical neutrality. The internal polarization charge, Q_{pol}, is indicated at the top of each sequence and the associated polarization field lines are also shown. The conduction charge, Q_{cond} which is transported between the present and previous sequence is also indicated at the bottom of sequences in which charge transport has occurred. Sequence (c') illustrates electron trapping into a deep level giving rise to fixed charge; note that the consequence of this trapping is a reduction of Q_{cond}, as shown between (c') and (d), and a reduction in the polarization charge of the subsequent opposite polarity pulse, as indicated in sequence (d). Electron trapping at the other interface is illustrated in sequence (d'). As can be seen by comparing sequences (b) and (c) to (e) and (f), the net effect of trapping electrons in deep levels is to decrease Q_{cond} by an amount equal to the amount of charge trapped and to decrease Q_{pol} by an amount equal to one-half of the charge trapped. This reduction in Q_{pol} leads to a reduction in the polarization field, which aids the subsequent voltage pulse, with a concomitant increase in the turn-on voltage.

B. Physical mechanisms of aging

As found in the previous section, ACTFEL aging appears to be due to electron trapping at deep levels located at or very near the interfaces. These deep levels are not dischargeable by the electric field so they contribute to the fixed charge in contradistinction to interface state charge. What is the physical nature of these deep, fixed charge traps and how do they arise?

With respect to the second part of the above question, how do the deep, fixed charge traps responsible for ACTFEL aging arise, there seem to be two possible answers. First, these deep traps may exist in the initially fabricated ACTFEL device but the capture probability may be very small and, thus, a long aging time may be required before these traps are filled with electrons. The capture probability is particularly small if the electrons to be captured have to surmount an energy barrier, E_B, or tunnel a distance, x, into the interface to be captured. In such a case the capture cross section, σ, is exponentially reduced in comparison to the normal capture cross section, σ_o, as follows:

$$\sigma = \sigma_o \exp\left(\frac{-E_B}{k_b T}\right)\exp\left(\frac{x}{x_t}\right), \tag{10}$$

where x_t is a characteristic tunneling distance. This type of aging instability, which is subsequently denoted as charge injection and trapping, arises in a wide variety of metal-insulator-semiconductor (MIS) systems including amorphous silicon,[23] InP,[24] and CdSe.[25] Charge injection and trapping instabilities are characterized by logarithmic aging kinetics and a very weak temperature dependence;[23] the aging activation energy has been reported to be 0.04–0.05 eV for InP MIS capacitors[24] and 0.12 eV for CdSe thin-film transistors.[25]

Although the logarithmic aging kinetics associated with charge injection and trapping are consistent with the experimental short-term aging data, the activation energy is a bit too large to be consistent with this mechanism. Furthermore, recent experiments provide additional evidence that ACTFEL aging does not occur by charge injection and trapping. In these experiments, attempts were made to reset the turn-on voltage after aging by photodepopulation of deep traps. Three optical sources, a tungsten lamp, a high-pressure mercury lamp, and a nitrogen laser, were employed in the infrared, visible, and ultraviolet regions of the electromagnetic spectrum for photodepopulation. Optical irradiation of the ACTFEL device was accomplished under open- and short-circuit conditions for various irradiation times. These experiments yielded no evidence for turn-on voltage resetting by photodepopulation. Since the turn-on voltage could not be reset optically, it is concluded that the aging is irreversible, which is inconsistent with charge injection and trapping in which the turn-on voltage can be reversibly reset.[24]

The second possible way in which the deep traps responsible for aging instabilities arise is that they are generated by atomic rearrangement at the interface. This is the mechanism we prefer to invoke to explain the experimentally observed ACTFEL aging trends. The aging is envisaged to occur as follows. Energetic hot electrons impinge against the insulator conduction band discontinuity and must dissipate a significant amount of energy, denoted E_{diss}, to thermalize to the bottom of the phosphor conduction band and are subsequently trapped at interface states. If the aging mechanism involves atomic migration, it is most likely that the hot electron thermalization energy, E_{diss}, is of primary importance in facilitating bond breakage and atomic rearrangement in the impact zone located near the interface. The activated atomic species can then migrate away from the impact zone into either the SiON or the ZnS under the influence of electric field and temperature.

If such atomic migration is responsible for ACTFEL aging the experimentally deduced aging activation energy of 0.2 eV must be identified as the migration energy, ΔH_m, for atomic migration. ΔH_m's for nearest- and second-nearest neighbor hopping into a vacancy are estimated using Van Vechten's ballistic model for atomic migration[26]

$$\Delta H_m = \tfrac{1}{2} m (F d k_B \theta_d / h)^2, \tag{11}$$

where m is the mass of the atom which hops, F is a geometric constant equal to 0.9, d is the distance which the atom hops, k_B is Boltzmann's constant, θ_d is the Debye temperature, and h is Planck's constant. For the purpose of these estimates, the hopping distances are assumed to be the equilibrium nearest- and second-nearest neighbor

TABLE I. Estimated migration energies for nearest- and second-nearest neighbor hopping.

Hopping atom	ΔH_m(eV) Nearest-neighbor	ΔH_m (eV) Second-nearest neighbor
Zn	0.81	2.16
S	0.40	1.05
Si	0.35	0.53
O	0.20	0.94
N	0.17	0.47

atomic distances in ZnS(i.e., 2.35 and 3.38 Å) and the Debye temperature of ZnS ($\theta_d = 350$ K) (Ref. 27) is assumed. The estimated values of ΔH_m are collected in Table I. Note that these estimates are not expected to be precise because of inherent limitations of the ballistic model and also because of the questionable validity of assuming bulk ZnS properties at the interface. With these limitations in mind, an analysis of Table I suggests that the experimentally observed activation energy of 0.2 eV is most compatible with interdiffusion by nearest-neighbor hopping at the interface, whereas second-nearest neighbor hopping would seem to be precluded because of the large predicted ΔH_m's. It should be noted that vacancy self-diffusion in a compound semiconductor such as ZnS is a conceptually simple process if it occurs by second-nearest neighbor hopping since it occurs only on a single sublattice. In contrast, nearest-neighbor hopping in ZnS is a much more complicated self-diffusion process in which hopping transforms a simple vacancy into a more complicated vacancy-antisite defect. Thus, nearest-neighbor hopping naturally gives rise to the creation of defect complexes which could be responsible for the fixed charge, deep levels responsible for aging instabilities. For example, suppose that sulfur vacancies, V_S's exist at the SiON/ZnS interface, as has been previously suggested by various researchers.[7,11,12,14] A possible atomic mechanism for aging instabilities could involve O or N diffusion by nearest-neighbor hopping into a sulfur vacancy which could be described by the following defect reactions:

$$O_O + V_S \underset{nnh}{\rightarrow} O_S + V_O, \tag{12}$$

$$N_N + V_S \underset{nnh}{\rightarrow} N_S + V_N. \tag{13}$$

The defect complex specified by the right side of the above defect reactions would presumably be the source of the fixed charge, deep levels responsible for ACTFEL aging. Such a process of electron trapping due to deep level formation by nearest-neighbor hopping has been previously invoked[24] to explain threshold voltage instabilities in InP MIS capacitors.

The proposed atomic scenario presumes that the required interdiffusion occurs within atomic distances of the interface. The viability of this assertion can be tested by noting that the maximum amount of trapped charge is equal to $(\Delta Q_{cond})^{max}$, which from Fig. 11(b) is about 31 nC. If each atomic defect captures only one electron and these defects exist within 10 Å of the interface, this yields a density of atomic defects of order $10^{19}/cm^3$. (The device area is 0.079 cm^2.) Thus, a 0.1% concentration of sulfur vacancies initially residing within 10 Å of the interface is sufficient to give rise to the aging instabilities observed experimentally; this appears to be a viable possibility.

V. CONCLUSIONS

An ACTFEL aging study is presented in which the *C-V* technique is used to characterize the electrical properties of evaporated ZnS:Mn ACTFEL devices at various temperatures as a function of aging time. The primary experimental findings of this study are the following: (1) The insulator and phosphor capacitances are constant with respect to aging time; this suggests that the perturbation in the electrostatic charge distribution which is responsible for aging occurs near the SiON/ZnS interfaces. (2) The *C-V* curve shifts rigidly with aging time; a rigid shift indicates that changes in the fixed charge density, not the interface charge density, give rise to aging. (3) The *C-V* turn-on voltage increases while the polarization and conduction charges decrease as a function of aging time. Also, the internal phosphor turn-on field increases slightly with respect to aging time. These experimental observations imply that conduction charge originally available for transport is trapped in deep traps at the SiON/ZnS interface, reducing the polarization charge and, thus, increasing the turn-on voltage. (4) The activation energy for short-term aging is found to be approximately 0.2 eV.

These experimental observations lead to the following model for aging of evaporated ZnS:Mn ACTFEL devices. Atomic rearrangement at SiON/ZnS interfaces leads to the formation of deep level, fixed charge states which trap transported conduction electrons. Such electron trapping leads to a reduction of the conduction and polarization charges and an increase in the turn-on voltage. It is likely that atomic migration is stimulated by the thermalization energy dissipated by hot electrons after they impinge upon the SiON conduction band discontinuity. Also, it is likely that atomic migration at the interface is exacerbated by the presence of sulfur vacancies in the ZnS near the interface.

ACKNOWLEDGMENTS

We wish to thank Ron Khormaei, Chris King, Dick Coovert, Dick Tuenge, Sey-Shing Sun, Eric Bringuier, and John Arthur for useful discussions during the course of this work. This work was supported by the Air Force Office of Scientific Research under contract AFOSR-89-0309 and by the U. S. Army Research Office under contract DAAL03-91G0242.

[1] T. Inoguchi, M. Takeda, Y. Yakihara, Y. Nakata, and M. Yoshida, SID 74 Dig. 5, p. 84 (1974).
[2] T. Inoguchi and S. Mito, *Topics in Applied Physics* (Springer, Heidelberg, 1977).
[3] K. Okamoto, Y. Nasu, and Y. Hamakawa, IEEE Trans. Electron Devices **ED-28**, 698 (1981).
[4] K. W. Yang, Ph. D. thesis, Oregon State University, 1981.

[5] P. M. Alt, D. B. Dove, and W. E. Howard, J. Appl. Phys. **53**, 5168 (1982).
[6] K. W. Yang and S. J. T. Owen, IEEE Trans. Electron Devices **ED-30**, 452 (1983).
[7] K. Taniguchi, K. Tanaka, T. Ogura, Y. Kakihara, S. Nakajima, and T. Inoguchi, 1985 SID Proc. **16**, p. 231 (1985).
[8] G. O. Müller, R. Mach, R. Reetz, and G. U. Reinsperger, SID 88 Dig. **19**, p. 23 (1988).
[9] M. Nishikawa, T. Matsuoka, T. Tohda, Y. Fujita, J. Kuwata, and A. Abe, SID 88 Dig. **19**, p. 19 (1988).
[10] J. Watanabe, M. Wakitani, S. Sato, and S. Miura, SID 88 Dig. **19**, p. 288 (1988).
[11] I. Khormaei, M. S. thesis, Oregon State University, 1989.
[12] R. Khormaei, J. F. Wager, and C. N. King, SID 89 Dig. **20**, p. 65 (1989).
[13] R. Mach and G. O. Mueller, *Electroluminescence* (Springer, Berlin, 1989).
[14] R. Khormaei, C. N. King, R. E. Coovert, and J. F. Wager, SID 91 Dig. **22**, p. 74 (1991).
[15] Y. H. Lee, I. J. Chung, and M. H. Oh, Appl. Phys. Lett. **58**, 962 (1991).
[16] R. C. McArthur, J. D. Davidson, J. F. Wager, I. Khormaei, and C. N. King, Appl. Phys. Lett. **56**, 1889 (1990).
[17] J. D. Davidson, J. F. Wager, I. Khormaei, C. N. King, and R. Williams, IEEE Trans. Electron Devices (to be published).
[18] J. D. Davidson, M. S. thesis, Oregon State University, 1991.
[19] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
[20] E. Bringuier, J. Appl. Phys. **66**, 1314 (1989).
[21] Y. A. Ono, H. Kawakami, M. Fuyama, and K. Onisawa, Jpn. J. Appl. Phys. **26**, 1482 (1987).
[22] R. Mach and G. O. Müller, Phys. Status Solidi A **81**, 609 (1984).
[23] M. J. Powell, C. van Berkel, and J. R. Hughes, Appl. Phys. Lett. **54**, 1323 (1989).
[24] M. T. Juang, J. F. Wager, and J. A. Van Vechten, J. Electrochem. Soc. **135**, 2019, 2024 (1988).
[25] S. W. Wright and J. C. Anderson, Thin Solid Films, **62**, 89 (1979).
[26] J. A. Van Vechten, Phys. Rev. B **12** 1247 (1975).
[27] R. R. Reeber, Phys. Status Solidi A **26**, 253 (1974).
[28] J. A. Van Vechten, J. Appl. Phys. **53**, 7082 (1982).

ELECTRICAL CHARACTERIZATION AND MODELING OF ACTFEL DEVICES

J.F. Wager and A.A. Douglas, Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, OR 97331-3211
D.C. Morton, Electronics Technology and Devices Laboratory, U.S. Army LABCOM, Ft. Monmouth, NJ 07733

ABSTRACT

A brief review is provided summarizing charge-voltage (Q-V) and capacitance-voltage (C-V) characterization of alternating-current thin-film electroluminescent (ACTFEL) devices. A synopsis of SPICE modeling of ACTFEL devices as related to electrical characterization is also included.

1. INTRODUCTION

The purpose of this paper is to review certain electrical methods used to characterize alternating-current thin-film electroluminescent (ACTFEL) devices. Primary emphasis is placed on the charge-voltage (Q-V) and capacitance-voltage (C-V) techniques. Also, SPICE modeling as relevant to the electrical characterization of ACTFEL devices is briefly discussed. Much of what is contained in this paper originates from previously published work.[1-17]

2. EXPERIMENTAL TECHNIQUE

A. Measurement Set Up

The basic generic circuit used for the electrical characterization of ACTFEL devices is shown in Fig. 1 where "series" and "sense" refer to locations where possible circuit elements or probes are located and "D.U.T." stands for device under test. $v_1(t)$, $v_2(t)$, and $v_3(t)$ refer to the three locations at which the voltage can be monitored as a function of time.

Table 1 summarizes five circuit configurations which have been used for the electrical characterization of ACTFEL devices. Note that the purpose of these circuits is to assess the two-terminal ACTFEL characteristics by measuring $v_1(t)$, $v_2(t)$, $v_3(t)$, or the total current, i(t). The fifth column of Table 1 entitled "i(t) Expression" refers to the relationship used to evaluate i(t) in terms of more fundamental, measurable quantities.

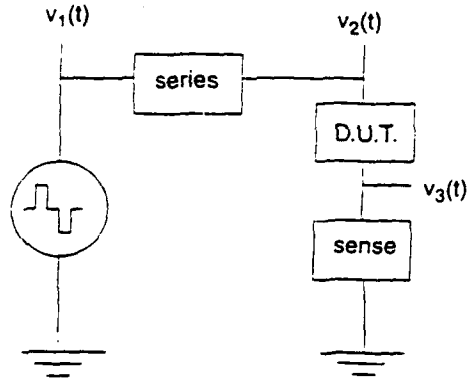


Figure 1. Basic generic circuit for ACTFEL characterization.

The main advantage of methods 1 and 2 is their very good time resolution. In contrast, the RC time constant associated with methods 3-5 smears out the measured current or voltage transient, thus, obscuring the dynamic response at short time. However, methods 3-5 are advantageous for experiments such as long-term aging studies[17] in which the short-time response is of less significance than the lifetime of the device; the series resistor acts as a current-limiter, protecting the ACTFEL from catastrophic breakdown.

B. Assessment of Electrical Quantities

The two fundamental dynamic quantities of importance for ACTFEL electrical characterization are total current, i(t), and the total voltage across the ACTFEL device, $v_{el}(t)$. i(t) is obtained as indicated in Table 1 whereas $v_{el}(t) = v_2(t) - v_3(t)$. From these fundamental, measurable quantities several other electrical quantities may be deduced as follows.

Table 1. Circuit configurations used for ACTFEL electrical characterization.

Method Number	Series Device	Sense Device	Circuit Designation	i(t) Expression	References
1	—	capacitor (C_s)	Sawyer-Tower	$C_s \frac{dv_3}{dt}$	2,9
2	current probe	—	current probe method	measured directly	14
3	resistor (R_s)	—	series resistor method	$\frac{v_2(t) - v_1(t)}{R_s}$	12
4	resistor (R_s)	resistor (R_c)	current sense resistor method	$\frac{v_3(t)}{R_c}$	15-17
5	resistor (R_s)	capacitor (C_s)	resistor/capacitor method	$\frac{v_2(t) - v_1(t)}{R_s}$	8

For methods 2-4 the instantaneous charge, q(t), is evaluated as the integral of the current,

$$q(t) = \int_{\tau=0}^{t} i(\tau)d\tau \ . \tag{1}$$

Note that q(t) is given directly by $q(t) = C_s v_3(t)$ for methods 1 and 5 which employ sense capacitors. In these cases i(t) can be calculated as,

$$i(t) = \frac{dq(t)}{dt} \ , \tag{2}$$

or, alternatively, by the expressions shown in Table 1. A Q-V curve is obtained by plotting q(t) versus $v_{el}(t)$. It is important to note that q(t) evaluated using methods 2-4, which do not utilize a sense capacitor, is arbitrary to within an integration constant so that the q(t) and Q-V offsets are unspecified. This Q-V offset problem can be overcome by acquiring i(t) over a full period of the waveform. The following condition,

$$\int_{t=0}^{T} q(t)dt = 0 \ , \tag{3}$$

where T is the voltage waveform period, can then be imposed upon the q(t) curve calculated from i(t) to obtain the proper offset. The condition specified by (3) arises from steady-state charge conservation. In contrast, methods 1 and 5, which employ sense capacitors, give Q-V curves with appropriate offsets along the y-axis since the sense capacitor accounts for the integration constant responsible for this offset.

The dynamic capacitance is equal to the first derivative of the charge with respect to the voltage across the ACTFEL device,

$$C\left[v_{el}(t)\right] = \frac{dq(t)}{dv_{el}(t)} \tag{4}$$

or, equivalently, as the total current divided by the time derivative of the voltage across the ACTFEL device,

$$C\left[v_{el}(t)\right] = \frac{i(t)}{\frac{dv_{el}(t)}{dt}} \ . \tag{5}$$

The C-V curve is found by plotting $C[v_{el}(t)]$ versus $v_{el}(t)$.

The electric field in the phosphor, $\xi_p(t)$, is given by,[11]

$$\xi_p(t) = \frac{1}{d_p}\left[v_{el}(t) - \frac{q(t)}{C_i}\right] \tag{6}$$

where d_p is the thickness of the phosphor and C_i is the total insulator capacitance of the ACTFEL device. The validity of (6) depends upon the absence of bulk charge in the phosphor so that the field is constant across the phosphor.

The conduction current, $i_{cond}(t)$ is usually found using a capacitive bridge technique[1,5,7,10,13,14] in which the displacement current is nulled out below threshold.

Alternatively, if it is again assumed that there is no bulk space charge, $i_{cond}(t)$ may be calculated[1] as,

$$i_{cond}(t) = \frac{C_i + C_p}{C_i}\, i(t) - C_p \frac{dv_{el}(t)}{dt} . \tag{7}$$

Plotting $i_{cond}(t)$ versus $\xi_p(t)$ yields the current-field (I-ξ) curve. The advantage of I-ξ plots is that they graphically display the nonlinear resistor characteristics which are important for establishing the device physics of ACTFEL performance. The primary disadvantage of this technique is that its accuracy depends on the precise establishment of C_i.[13]

3. SPICE SIMULATION OF ACTFEL DEVICES

Perhaps the simplest equivalent circuit[2] for an ACTFEL device is indicated in Fig. 2. C_{i1} and C_{i2} are insulator capacitances and C_p is the phosphor capacitance. The phosphor capacitance is shunted by two back-to-back Zener diodes which account for conduction and field clamping in the phosphor. SPICE simulation using this simple equivalent circuit describes the main features of ACTFEL operation but does not account for the detailed electrical characteristics measured experimentally.

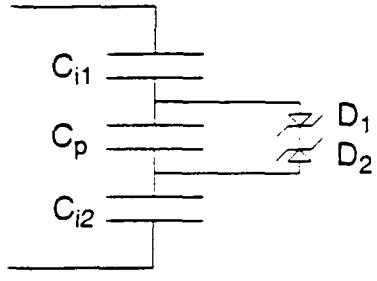


Figure 2. Simple ACTFEL equivalent circuit.

An improved equivalent circuit[15,16] which accounts for more details in the two-terminal characteristics exhibited by ACTFEL devices is given in Fig. 3. R_{i1}, R_{i2}, and R_p represent shunt resistances of the respective insulators or phosphor. R_{ito} accounts for the non-zero series resistance of the conducting, transparent electrode. R_d is denoted the diode resistance.

Currently, our most complete ACTFEL model[18] consists of inserting a parallel resistor-capacitor (RC) combination between R_d and the lower node. This RC parallel network more precisely accounts for the dynamic behavior observed in i(t) or $\xi_p(t)$ characteristics.

4. ACTFEL ELECTRICAL CHARACTERIZATION METHODS

A. Charge-Voltage (Q-V) Analysis

i. Ideal Q-V Characteristics

A set of ideal Q-V curves is indicated in Fig. 4. Curve (a) corresponds to a situation in which the maximum voltage across the ACTFEL device, V_m, is less than the threshold voltage, V_{th}. V_{th} corresponds to the critical applied voltage across the ACTFEL device at which appreciable conduction current begins to flow. Since $V_m < V_{th}$ for curve (a), the Q-V curve is linear, as expected for a capacitor.

Curve (b) in Fig. 4 corresponds to $V_m \geq V_{th}$ such that an appreciable amount of conduction current begins to flow with a concomitant initiation of hysteresis. It is important to note that Q-V curves are obtained for a constant V_m and that V_{th} corresponds to the critical V_m at which conduction and Q-V hysteresis are initiated. Thus, Q-V curves obtained at constant V_m are distinctly different in how they are obtained compared to brightness-voltage (B-V_m) curves which are generated by varying V_m. It should also be noted that the

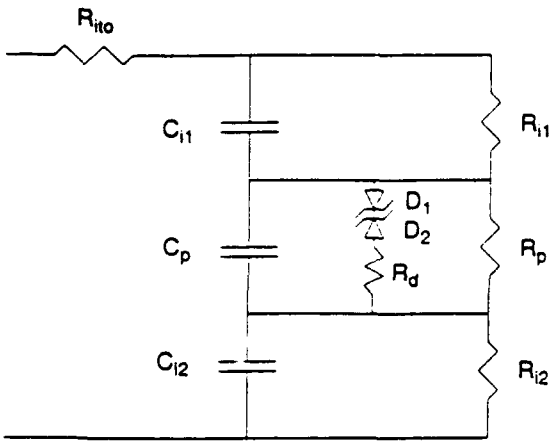


Figure 3. Improved ACTFEL equivalent circuit model for SPICE simulation.

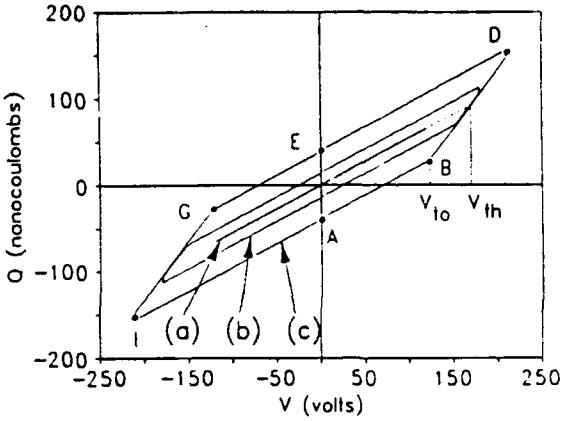


Figure 4. Ideal Q-V curves.

ACTFEL threshold is not distinct (i_{cond} depends approximately exponentially on ξ_p) and, thus, V_{th} is not precisely defined. In this respect, ACTFEL threshold characteristics are analogous to subthreshold effects in short-channel MOSFETs.[19]

Returning to Fig. 4, curve (c) represents a more typical Q-V curve where $V_m > V_{th}$. The labels A, B, D, E, G, and I are used to designate specific points on the Q-V curve and are used in the subsequent discussion. V_{th} is the value of V_m at which conduction current begins to flow. When V_m exceeds V_{th}, conduction current causes an accumulation of polarization charge at the phosphor/insulator interface which assists conduction on the subsequent pulse. Thus, V_{to}, the turn-on voltage, is lowered below V_{th} and conduction in initiated when $V_{cl}(t) = V_{to}$.

Again returning to Fig. 4, curve (c), the total capacitance, C_t, of the ACTFEL is obtained from the slope of the Q-V curve in the pre-turn-on regime. Similarly, the total insulator capacitance, C_i, is given by the slope of the Q-V curve above turn-on. Hence, the phosphor capacitance, C_p, can be calculated from

$$\frac{1}{C_p} = \frac{1}{C_t} - \frac{1}{C_i} . \tag{8}$$

ii. Realistic Q-V Characteristics

A realistic Q-V curve (i.e., a piecewise-linear curve showing the important features present in a real Q-V curve) is shown in Fig. 5 and differs from the ideal curve of Fig. 4 at $V = V_m^-$, V_m^+, and 0 V where the Q-V curve becomes essentially a multivalued function of voltage. In order to scrutinize Fig. 5 more closely, we supplement the labels of Fig. 4 with the additional labels, C, F, H, and J. Table 2 and Fig. 6 help clarify new features in the actual Q-V curve which are not present in the ideal curves.

Q^+_{cond} indicated in Fig. 5 is the total conduction charge transported across the phosphor during the positive voltage pulse. Note that for an asymmetrical ACTFEL device it is necessary to distinguish between charges arising from each voltage polarity, which is readily accomplished with the superscripts + and -. Q^{e+}_{pol} denotes the external (thus, the superscript "e") polarization charge stored at the phosphor insulator interface after the positive voltage pulse goes to zero. (No "e" superscript is required for Q_{cond} since when conduction charge flows the phosphor capacitance is shunted and the measured conduction current or charge is identical to that which flows internally in the ACTFEL.) Note that our definition of Q^e_{pol} differs from that of most authors who define polarization charge as the total difference in charge obtained from the Q-V curve at zero voltage. We prefer the present method since it enables us to handle asymmetrical devices in a more straightforward manner. Finally, the internal polarization charge, Q_{pol}, is related to Q^e_{pol} as follows:[7,9,11]

$$Q_{pol} = \frac{C_i + C_p}{C_i} Q^e_{pol} . \tag{9}$$

If field-clamping is assumed, it can be shown from a geometrical analysis of Fig. 5 that

$$V_{to}^{+} = \frac{Q_{pol}^{-}}{C_i} + V_{th} \qquad (10)$$

where Q_{pol}^{-} is a negative quantity.

Sections of the Q-V curve labeled EF and JA define Q_{leak}^{e+} and Q_{leak}^{e-}, the external leakage charge associated with the positive and negative applied voltage pulses, respectively. Q_{leak} arises from electron emission from shallow interface states during the zero voltage portion of the waveform, as clarified by Fig. 7. Notice that Q_{leak} reduces the polarization charge and therefore increases V_{to} as predicted by eqn. (10). Additionally, it is possible to measure the interface state density from $Q_{leak}(t)$ using a field-control circuit to adjust ξ_p between pulses.[20]

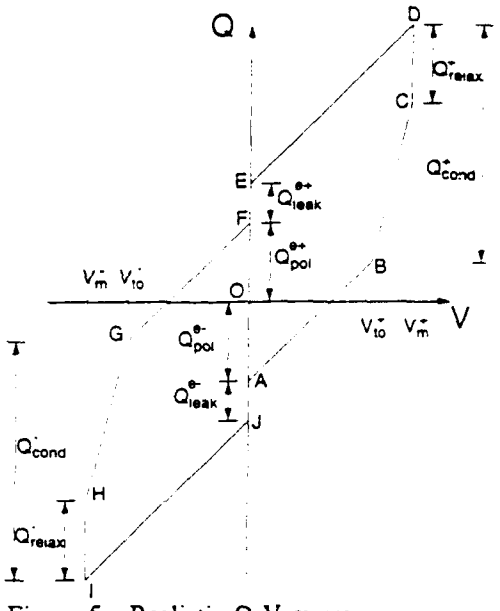


Figure 5. Realistic Q-V curve.

Sections CD and HI of Fig. 5 correspond to the portion of the applied voltage waveform where the voltage is constant and at its maximum value, as seen in Fig. 6(a). Point C is the time at which i(t) maximizes. Thus, CD corresponds to a region of the Q-V curve in which the conduction current tail flows and the phosphor field relaxes, thus defining the relaxation charge, Q_{relax}^{+}. For many applied voltage waveforms and measurement configurations, (e.g., large V_m, large pulse width, and large RC time constant) the Q_{relax}^{+} portion of the Q-V curve corresponds to a significant fraction of Q_{cond}^{+}.

Table 2. Q-V curve classification for positive pulse portions of the curve.

Section of Q-V Curve	Q-V Curve Classification	Comments
AB (FG)	prebreakdown	slope = C_{tot}
B (G)	V_{to}^{+}	turn-on voltage
BC (GH)	postbreakdown, large amount of conduction current flows	slope = C_i
BD (GI)	defines Q_{cond}^{+}	conduction charge
C (H)	i(t) maximizes	see Fig. 6
CD (HI)	postbreakdown, small amount of conduction current flows; defines Q_{relax}^{+}	see Fig. 6 relaxation charge
D (I)	V_m^{+}	maximum applied voltage
DE (IJ)	external voltage returns to zero	see Fig. 6
EF (JA)	defines Q_{leak}^{e+}	external leakage charge
OF (OA)	defines Q_{pol}^{e+}	external polarization charge

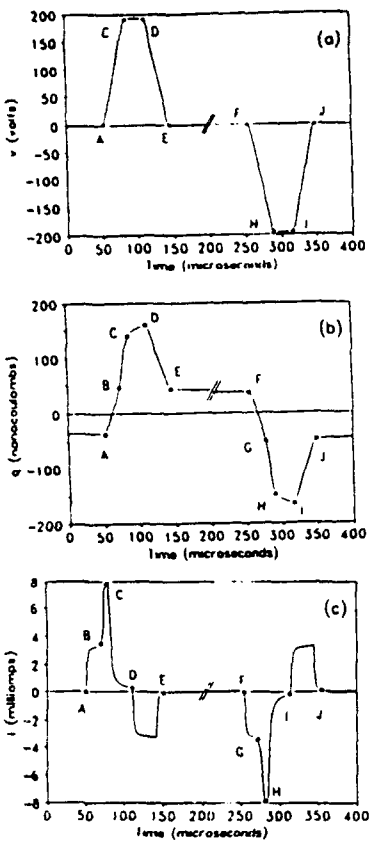


Figure 6. $v_{EL}(t)$, q(t), and i(t) curves.

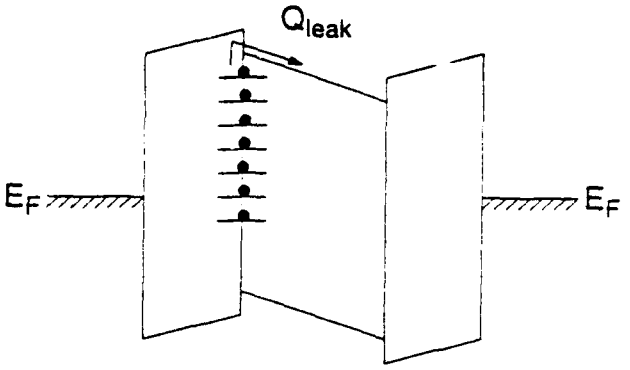


Figure 7. Energy band diagram illustrating the mechanism of leakage charge.

iii. Q-V_m Characteristics

The Q-V curves discussed in the previous section arise from plotting the instantaneous functions q(t) versus $v_{el}(t)$ while V_m is held constant during the measurement. An alternative approach to Q-V analysis is to measure i(t) as a function of V_m and to process i(t), as described in the following, in order to obtain a steady-state Q-V curve. We denote the steady-state Q-V curve as Q-V_m to distinguish it from the conventional instantaneous Q-V curve.

To obtain the Q-V_m curve first measure i(t) as a function of V_m from below threshold to about 30 V above threshold in 1 V steps. Then define t_1 and t_2 as shown in Fig. 8 with $t_2 - t_1 \simeq t_1$. Next define:

$$Q_{on}(V_m) = \int_{\tau=0}^{t_1} i(\tau)d\tau \tag{11}$$

and

$$Q_{off}(V_m) = \int_{\tau=t_1}^{t_2} i(\tau)d\tau \ . \tag{12}$$

Q_{off} corresponds to the displacement charge due to the falling portion of the applied voltage waveform while Q_{on} is the sum of the displacement and conduction charge of the previous portion of the waveform.

Plotting Q_{off} and Q_{on} versus V_m results in the Q-V_m curve is illustrated in Fig. 8(c). The primary advantage of this technique is that the same threshold voltage, V_{th}, as found from a B-V_m measurement is deduced. Additionally, the slope of Q_{on} and Q_{off} yield, respectively, C_i and C_t. However, C_i measured using the steady-state Q-V_m technique is a static capacitance which, in general, differs from C_i found from the dynamic Q-V method. The difference in the static and dynamic values for C_i arises from the relaxation charge. Note that C_i measured by the Q-V_m method corresponds to the capacitance to be charged by the voltage drivers whereas C_i measured by the Q-V technique is the dynamic capacitance relevant for assessment of the physical capacitance of the dielectric layers.

Returning to Fig. 8c, observe that the luminance or brightness, B, is plotted as a function of V_m. V_{th} is evident from the discontinuity in the B-V_m slope. The subthreshold, short-channel-like[19] effects mentioned previously are evident in this curve as B depends exponentially on V_m in the subthreshold regime.

B. Capacitance-Voltage (C-V) Analysis

One way of viewing C-V analysis is that the dynamic capacitance, i.e., the instantaneous current divided by the derivative of the voltage across the ACTFEL device with respect to time as indicated by eqn. (5), is

plotted as a function of the instantaneous voltage dropped across the ACTFEL device. This is the perspective from which we have viewed the C-V technique previously.[12,15-17] An alternative, and equivalent, viewpoint is that the C-V characteristic is simply the derivative of the Q-V curve, as implied by eqn. (4). Viewed from this perspective, it may be difficult to discern any advantage of the C-V technique over the Q-V method. There are several distinct advantages. It is our opinion, however, that these techniques are complementary, rather than rivals.

A typical C-V curve is shown in Fig. 9. C_i and C_t are defined by the flat portions of the C-V curve above and below the turn-on voltage, respectively. It is evident from Fig. 9 that turn-on is non-abrupt so that we define three turn-on voltages, V_{to1}, V_{to2}, and V_{to3}, which denote the onset, midpoint, and saturation of the C-V transition. SPICE simulation confirms that V_{to1} corresponds to the onset of conduction and V_{to3} to the initiation of field-clamping. V_{to2} closely corresponds to V_{to} assessed from Q-V analysis. The rising portion of the C-V curve beginning at approximately 190 V corresponds to the relaxation charge of the Q-V curve where $(dV_{el}(t))/dt = 0$ (i.e., CD and HI in Fig. 5); thus, the dynamic C-V technique provides no information during the relaxation charge portion of the waveform whereas the static C_i evaluated from Q-V_m technique includes relaxation charge contributions. Neither the static nor the dynamic C-V techniques (i.e., a static C-V curve may be obtained by differentiating Fig. 8c) provide any information regarding the leakage charge.

The slope of the C-V curve is inversely proportional to the density of interface states in the preclamping regime. A steep C-V slope implies a small preclamping interface state density, whereas a non-abrupt C-V transition indicates a high density. The preclamping interface state density, Q_{ss}, expressed in units of number of states per square centimeter is related to the C-V slope as follows:

$$Q_{ss} = \frac{C_i^2}{2qA}\,\frac{C_t}{C_p}\left[\frac{\Delta C}{\Delta V}\right]^{-1} \qquad (13)$$

where A is the ACTFEL device area and the term in the brackets is the slope of the C-V curve.

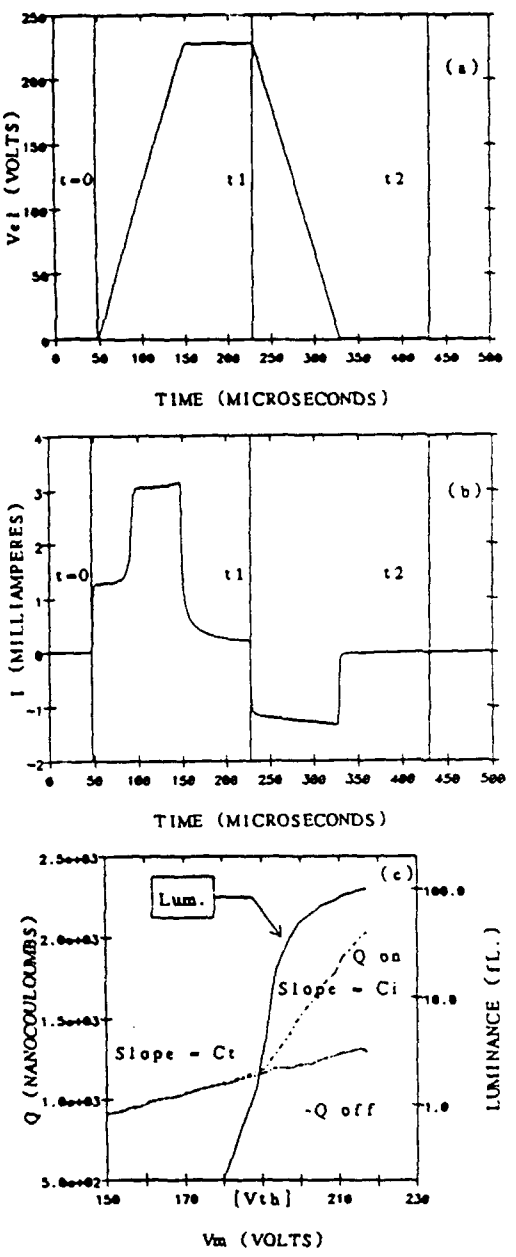


Figure 8. v_{el}, i(t), and Q-V_m curves used for steady-state charge-voltage assessment.

The existence of parasitic resistance in the ACTFEL can be inferred from the C-V curve, particularly in conjunction with SPICE modeling.[15,16] A SPICE simulated example of C-V curves exhibiting various ITO resistances (i.e., R_{ito} = 0, 30, and 100 Ω) is indicated in Fig. 10. Three distinct trends, labeled (1), (2), and (3), are the signatures by which a large R_{ito} parasitic resistance can be diagnosed from a C-V curve.

Perhaps the most useful aspect of a C-V curve is its immediate visual impact and the concomitant information it conveys. C_t and C_i can be evaluated at a glance, as can the relative magnitude of Q_{ss} and the existence of a large parasitic R_{ito}.

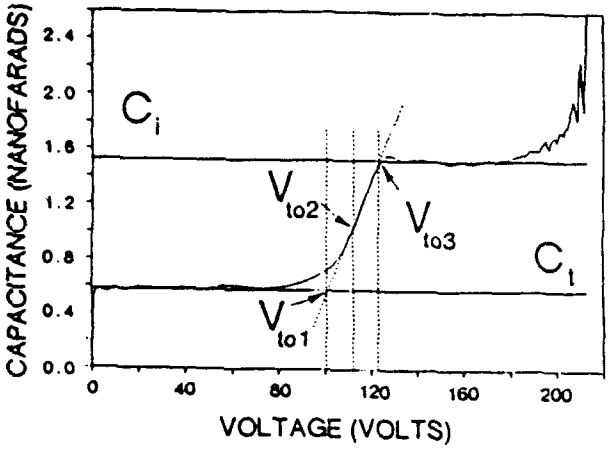


Figure 9. A typical C-V curve.

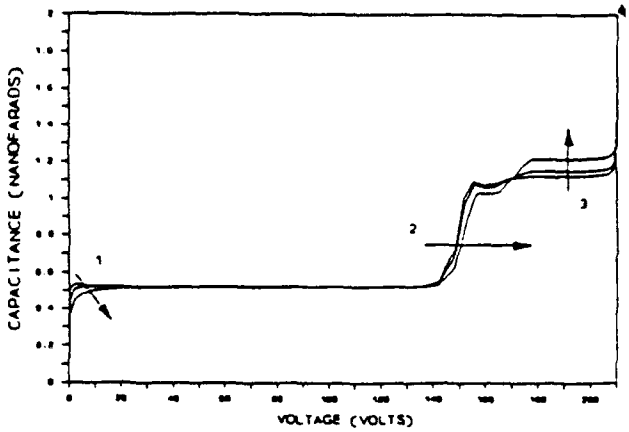


Figure 10. A SPICE-simulated C-V curve illustrating the effect of varying the large parasitic resistance, R_{ito}. R_{ito} = 0, 30, and 100 Ω. Arrows indicate an increasing value of R_{ito}.

5. EXAMPLES OF ACTFEL ELECTRICAL CHARACTERIZATION AND MODELING

A. ACTFEL Aging Studies

A study of the aging characteristics of evaporated ZnS:Mn devices as monitored by Q-V and C-V analysis has been recently reported.[15,17] Some of the primary experimental findings of this study are:

1. C_t and C_i are constant with respect to aging time.
2. The C-V curve shifts rigidly with aging time.
3. The C-V turn-on voltage increases while the polarization and conduction charges both decrease as a function of aging time.
4. The internal phosphor field at the onset of conduction increases slightly with aging time.

Observation 1 implies that the electrostatic charge redistribution responsible for the aging occurs near the phosphor/insulator interface. Observation 2 indicates that changes in the fixed charge density, not the interface state density, give rise to aging. Observations 3 and 4 imply that some of the original interface state charge is trapped in deep level, fixed charge states.

From these experimental observations an aging model was proposed as shown schematically in Fig. 11. The dissipation energy, E_{diss}, of hot electrons after they impinge upon the insulator conduction band discontinuity and thermalize to the phosphor conduction band minimum, initiates atomic rearrangement (indicated by the double arrows) at the phosphor/insulator interface. Atomic rearrangement leads to the creation of deep level, fixed charge states which trap electrons that would otherwise occupy interface states. Since electron emission cannot occur from these deep level, fixed charge states, the conduction and polarization charges are reduced as a function of aging time while the turn-on voltage and phosphor field increases.

This aging model could be deduced only after the electrostatics of the aging process were established from Q-V and C-V analysis. Furthermore, electrical characterization is readily computer automated which is a great advantage for aging studies which must be performed over very long durations.

B. Leakage Charge and SPICE Modeling

The phosphor resistance, R_p, illustrated in Fig. 3 is the parameter of prime importance in establishing the magnitude of the leakage charge, Q_{leak}, in a SPICE simulation. Two Q-V simulations are given in Fig. 12 using $R_p = 5 \times 10^5$ and 5×10^6 Ω. Note that for the larger value of R_p the Q-V curve exhibits Q_{leak} of virtually zero magnitude. The smaller value of R_p yields a simulated Q-V curve which exhibits trends more consistent with experiment.

From a SPICE point of view, Q_{leak} corresponds to a situation in which charge stored on plates of C_p discharges through R_p (the diode branch looks like an open circuit, refer to Fig. 4). Thus, Q_{leak} is modeled

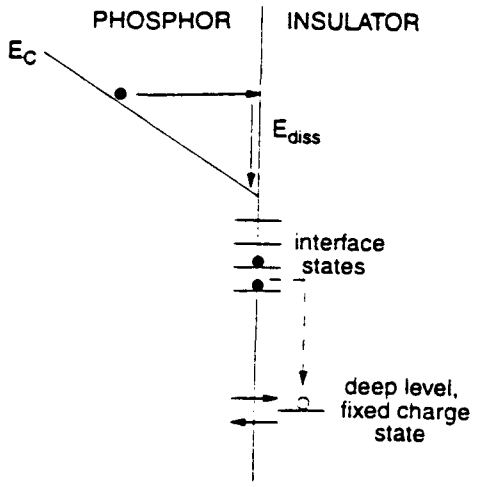


Figure 11. Proposed model for aging of evaporated ZnS:Mn ACTFEL devices.

in SPICE as a simple RC exponential decay where R_p and C_p are the relevant resistance and capacitance, respectively. The R_p C_p time constant is much greater than half of the waveform period for large R_p while they are almost equal for small R_p. Note that the present SPICE model predicts a simple exponential decay for $Q_{leak}(t)$ whereas experimentally it is usually found to be *nonexponential*.

C. SPICE Simulation of an ACTFEL Device with Asymmetric Interfaces

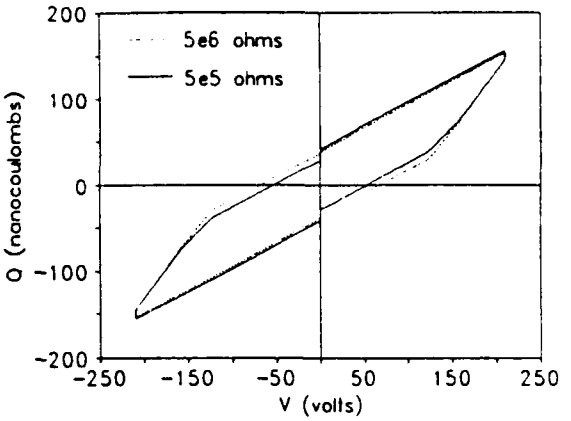


Figure 12. Two SPICE-simulated Q-V curves using $R_p = 5 \times 10^5$ and 5×10^6 Ω for the inner and outer curves, respectively.

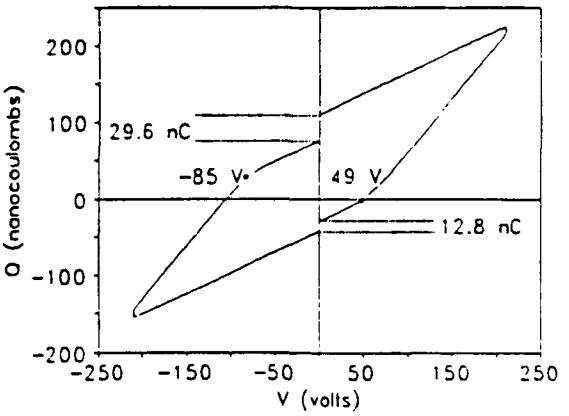


Figure 13. SPICE-simulated Q-V curve for an ACTFEL device with asymmetrical interfaces.

A SPICE Q-V simulation for an ACTFEL device with asymmetric interfaces is modeled by inequivalent Zener diode breakdown voltages of 50 and 100 V. This interface asymmetry yields $V^+_{to} = 49$ V and $V^-_{to} = -85$ V. Also, $Q^+_{cond} = 230$ nC compared to $Q^-_{cond} = 180$ nC and $Q^{e+}_{leak} = 29.6$ nC versus $Q^{e-}_{leak} = 12.8$ nC. Note that after steady-state is established in the ACTFEL, charge continuity requires that

$$Q^+_{cond} - Q^+_{leak} = Q^-_{cond} - Q^-_{leak} \,. \tag{14}$$

where Q^+_{leak} and Q^-_{leak} are the internal leakage charges given by

$$Q^+_{leak} = \frac{C_i + C_p}{C_i} Q^{e-}_{leak}$$

and (15)

$$Q^-_{leak} = \frac{C_i + C_p}{C_i} Q^{e-}_{leak} \,.$$

This charge continuity condition may be established with the aid of Fig. 13.

6. CONCLUSIONS

Q-V and C-V techniques are powerful, complementary methods for characterizing the two-terminal electrical properties of ACTFEL display devices. SPICE modeling can supplement these techniques so that the internal device physics of ACTFELs can be further clarified.

ACKNOWLEDGMENTS

We wish to thank Ahmad Abu-Dayah, Eric Bringuier, Jim Davidson, Ron Khormaei, Chris King, Shiro Kobayashi, Ralph McArthur, and Dick Williams for their many contributions to the work reviewed herein. This work was supported by the Air Force Office of Scientific Research under contract AFOSR-83-0309, the U.S. Army Research Office under contract D11L03-91G0242, and by the Electronic Technology and Devices Laboratory under contract DAAL01-89-C-0927 as administered by GEO-CENTERS, Inc.

REFERENCES

1. Y.S. Chen and D.C. Krupka, *J. Appl. Phys. 43*, 4089 (1972).
2. D.H. Smith, *J. Lumin. 23*, 209 (1981).
3. K.W. Yang, S.J.T. Owen, and D.H. Smith, *IEEE Trans. Electron Devices ED-28*, 703 (1981).
4. K.W. Yang, Ph.D. thesis, Oregon State University, 1981.
5. R. Mach and G.O. Müller, *Phys. Status Solidi A 69*, 11 (1982).
6. P.M. Alt, *Proc. SID25*, 123 (1984).
7. R. Mach and G.O. Müller, *Phys. Stat. Sol. A 81*, 609 (1984).
8. H. Onnagawa, M. Shibata, and K. Miyashita, *Jpn. J. Appl. Phys. 25*, 12 (1986).
9. Y.A. Ono, H. Kawakami, M. Fuyama, and K. Onisawa, *Jpn. J. Appl. Phys. 26*, 1482 (1987).
10. G.O. Müller, R. Mach, B. Selle, and G. Schulz, *Phys. Stat. Sol. A 110*, 657 (1988).
11. E. Bringuier, *J. Appl. Phys. 66*, 1314 (1989).
12. R.C. McArthur, J.D. Davidson, J.F. Wager, I. Khormaei, and C.N. King, *Appl. Phys. Lett. 56*, 1889 (1990).
13. K. Neyts and P. De Visschere, *Acta Polytechnica Scandinavica 170*, 291 (1990).
14. V.P. Singh, S. Krishna, and D.C. Morton, *J. Appl. Phys. 70*, 1811 (1991).
15. J.D. Davidson, M.S. Thesis, Oregon State University, 1991.
16. J.D. Davidson, J.F. Wager, I. Khormaei, C.N. King, and R. Williams, IEEE *Trans. Electron Devices ED-39* (1992).
17. J.D. Davidson, J.F. Wager, and S. Kobayashi, *J. Appl. Phys. 71* (1992).
18. A.A. Douglas and J.F. Wager (these proceedings).
19. S.M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
20. A.A. Douglas, J.F. Wager (to be published).

Transient decay of persistent photoconductivity in $Al_{0.3}Ga_{0.7}As$

T. W. Dobson
Tektronix, Inc., Beaverton, Oregon 97077

L. V. A. Scalvi[a] and J. F. Wager
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

(Received 23 June 1989; accepted for publication 29 March 1990)

Transient decay of persistent photoconductivity (TDPPC) measurements were performed and analyzed in terms of models in which the TDPPC is associated with thermally activated electron capture into DX and a modification of the ionized impurity density, and hence the mobility, concomitant with electron capture. Quantitative agreement between theory and experiment was possible when Chadi and Chang's model for DX [Phys. Rev. Lett. **61**, 873 (1988); Phys. Rev. B **39**, 10063 (1989)] was employed in conjunction with a photo-induced shallow donor.

I. INTRODUCTION

$Al_xGa_{1-x}As$ has attracted much attention due to properties which make it useful for many types of heterojunction device structures such as lasers, bipolar transistors, and field-effect transistors. An extremely unusual deep level in $Al_xGa_{1-x}As$, denoted DX, is known[1-11] to give rise to many metastable properties, the best known being persistent photoconductivity (PPC). Many studies have been undertaken in order to establish the basic physics of DX. These studies have yet to identify the atomic nature of the defect, let alone the fundamental physics. There is even a controversy as to whether the DX center is a donor-like or acceptor-like defect.

The debate over the donor- or acceptor-like nature of DX is due to conflicting photo-Hall measurements reported by various authors. Nelson[1] and Lang *et al.*[2] observed a decrease in the Hall mobility when the DX center was photoionized at low temperature, which they interpret to indicate that DX is donor-like. On the other hand, Saxena, Kunzel *et al.*, and Chand *et al.*[3,5,6,8] have reported the mobility to increase upon photoionization. They interpret this as evidence that DX is acceptor-like.

Part of the reason for the above discrepancy has been pointed out by Collins *et al.*[4] and confirmed by Nicholas *et al.*[7] Collins *et al.* performed photo-Hall measurements on two types of AlGaAs epitaxial structures. It was found that the increase or decrease of the Hall mobility upon photoionization of DX was a strong function of the multilayered nature of the AlGaAs sample. In particular, they asserted that photoinduced charge separation at the $Al_xGa_{1-x}As/GaAs$ interface leads to artifacts in photo-Hall measurements. Nicholas *et al.* confirmed this photoinduced charge separation and demonstrated that a two-dimensional electron gas (2DEG) is formed at the AlGaAs/GaAs interface. Chand *et al.*[6,8] were cognizant of the possibility of unintentionally forming a 2DEG and employed compositional grading to prevent its formation.

The purpose of this work is to propose an alternative technique, transient decay of persistent photoconductivity (TDPPC), for determining the donor- or acceptor-like nature of DX. We will show that the TDPPC measurements are qualitatively consistent with the negative-U, acceptor model for DX proposed by Chadi and Chang.[10,11] Quantitative agreement between the experimental TDPPC data and this model could only be obtained, however, by postulating the existence of an additional, photo-induced shallow donor; such a donor has recently been deduced by Jia *et al.*[12]

II. EXPERIMENTAL TECHNIQUE

The $Al_xGa_{1-x}As$ samples used in this work were grown by molecular beam epitaxy (MBE) on an undoped, semi-insulating GaAs substrate. All of the results reported in this paper refer to a 2-μm thick $Al_{0.3}Ga_{0.7}As$ n-type active layer doped with silicon ($N_{Si} = 1\times10^{18}$ cm^{-3}). Care was taken to avoid the formation of a 2DEG using the following sample structure: a 0.2-μm undoped GaAs buffer layer was grown on the semi-insulating substrate followed by a 500 Å buffer layer of undoped $Al_xGa_{1-x}As$ that is compositionally graded from $x = 0$ to $x = 0.3$, a 2 μm layer of undoped $Al_{0.3}Ga_{0.7}As$, and the 2 μ active layer.

Ohmic contacts were fabricated by evaporating AuGeNi dots through a shadow mask in a vacuum evaporator. The contacts were annealed at 400 °C in forming gas.

Measurements of the TDPPC were performed using an automated system based on Hewlett-Packard equipment consisting of a model 9336 desktop computer in conjunction with a 4280A 1 MHz capacitance meter. The temperature was controlled using an Air Products closed-cycle helium cryostat and a Scientific Instruments temperature controller. The light source consisted of a tungsten lamp, Jarrel-Ash monochrometer, and appropriate filters and lenses.

The TDPPC measurement is performed as follows. The sample is initially cooled in the dark to the desired measurement temperature of 80–100 K. The sample is then illuminated with sub-bandgap, monochromatic light until the photoconductance reaches steady state. The light source is

[a] Permanent address: Institute Fisica e Quimica de Sao Carlos-USP-13560-Sao Carlos-SP, Brazil.

 0021-8979/90/140601-05$03.00

turned off at $t = 0$ and the conductance transient is monitored with the capacitance meter.

The experimental results are then fit to theoretical curves based on models discussed below. The quality of the fit is used to judge the viability of the DX model.

An example of a TDPPC curve is shown in Fig. 1. A plot of the normalized inverse conductance versus time proved to be a better means of comparing the experimental results with the theoretical fit. Thus, all TDPPC data is plotted in this manner in the remainder of this paper.

III. TDPPC THEORY

TDPPC in $Al_xGa_{1-x}As$ with $x > 0.2$ can be explained in terms of the nonequilibrium capture kinetics of electrons from the conduction band into the DX center and the concomitant change in the mobility due to ionized impurity scattering. The theoretical TDPPC equation is given by

$$G(t) = qn(t)\mu(t), \tag{1}$$

where G, n, and μ denote the conductance, electron concentration in the conduction band, and the mobility, respectively.

The differential equation for the decay of photoionized electrons when the illumination is turned off is given by

$$\frac{dn}{dt} = -c_n[N_{DX}(\text{empty})] + (e_n^t + e_n^0)[N_{DX}(\text{occupied})], \tag{2}$$

where $N_{DX}(\text{empty})$ and $N_{DX}(\text{occupied})$ refer to whether the DX center is filled with an electron. Also, c_n, e_n^0, and e_n^t are the electron capture rate, electron optical emission rate, and electron thermal emission rate, respectively. Since the TDPPC transient occurs at low temperature and in the dark, $c_n \gg e_n^t, e_n^0$. The electron capture rate is given by

$$c_n = v_{th}\sigma_n n, \tag{3}$$

where v_{th} is the thermal velocity and σ_n is the electron capture cross section which in the case of DX is thermally activated

$$\sigma_n = \sigma_\infty \exp[-(E_{cap} - E_F)/k_BT]. \tag{4}$$

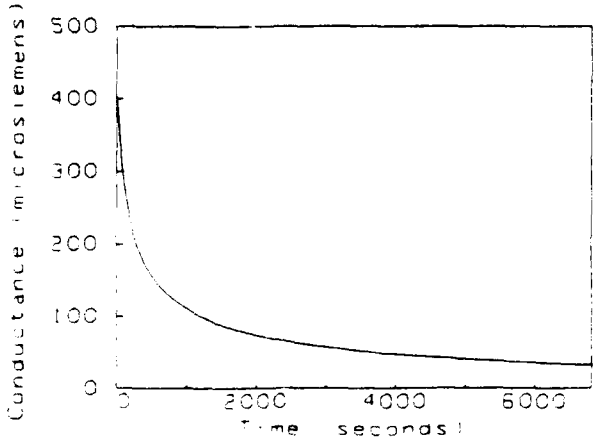


FIG. 1. TDPPC plot at 85 K after 1.37 eV photon illumination.

σ_∞ is the high-temperature capture cross section, E_{cap} is the capture barrier, E_F is the Fermi level, and k_B is Boltzmann's constant. E_F is obtained from a solution of

$$n = N_c F_{1/2}(\eta), \tag{5}$$

where

$$\eta = (E_F - E_C)/k_BT, \tag{6}$$

and where $F_{1/2}$ is the Fermi-Dirac integral of order 1/2 and N_C is the conduction band density of states as given by Casey and Panish.[13] It should be noted that the Fermi-level correction to the capture cross section in Eq. (4) is of great importance;[14] we previously concluded[15] (erroneously) that DX is donor-like because we neglected this correction.

At the temperatures used in this study (80–100 K) and with the high concentration of impurities incorporated into the sample ($N_{Si} = 1 \times 10^{18}\,cm^{-3}$), the mobility is dominated by ionized impurity scattering.[16] Since the capture of electrons into the DX center represents a change in the charge state of the dominant level, the mobility can change dramatically during the TDPPC; this change is expected to be different depending on whether DX is donor- or acceptor-like. We employ the Brooks-Herring equation for the mobility due to ionized impurity scattering, which is given by[17]

$$\mu_{ii} = \frac{128\sqrt{2\pi}\epsilon^2(k_BT)^{3/2}}{q^3 m_e^{1/2} N_{ii}\ln(b/n)}, \tag{7}$$

where

$$b = \frac{96\pi^2 m_e \epsilon (k_BT)^2}{q^2 h^2}. \tag{8}$$

where m_e is the electron conductivity effective mass and ϵ is the dielectric constant. Note that although we ignore scattering modes other than ionized impurity scattering, over the temperature range of this study this assumption amounts simply to the rescaling of the TDPPC curve whereas the basic shape of the curve will not be affected.

The TDPPC theory presented to this point [i.e., Eqs. (1)–(8)] is completely general, independent of how DX is modeled. To accomplish a TDPPC simulation, Eqs. (1)–(8) must be solved simultaneously. This can be achieved only after two parameters, $N_{DX}(\text{empty})$ and N_{ii}, are specified in terms of n. But $N_{DX}(\text{empty})$ and N_{ii} depend on the nature of DX and a model for DX must be postulated to make further progress in the TDPPC simulation.

A summary of the DX models considered herein is given in Table I. Note that $N_{DX}(\text{empty})$ arises from charge balance in conjunction with the constraint that AlGaAs be n-type; these conditions cannot be achieved simultaneously for the simple acceptor model. All of the models employ $n(0)$, the electron concentration in the conduction band at $t = 0$, as a freely adjustable parameter in the TDPPC simulation. In addition, a net concentration of shallow acceptors or donors (i.e., N_{SA} or N_{SD}) are used as adjustable parameters in models 3, 4, and 6. A third adjustable parameter, corresponding to the equilibrium concentration of ionized DX centers [i.e., $N_{DX}^+(\text{eq})$ and $N_{DX}^-(\text{eq})$ for DX assumed to be donor- or acceptor-like, respectively], is used in models 3 and 4. Note that $n(0)$ is freely adjustable for every temperature of interest whereas the other parameters are assumed to

TABLE I. A summary of the parameters used in the TDPPC simulation for each of the DX models considered. Notation: n, $n(0)$ = electron concentration in the conduction band at an arbitrary time and at $t = 0$. N_{SA}, N_{SD}, N_{Si} = total concentration of net shallow acceptors, net shallow donors, and incorporated Si atoms. $N_{DX}^{+}(eq)$, $N_{DX}^{-}(eq)$ = concentration of ionized DX in equilibrium assuming DX is donor- or acceptor-like.

Model	N_{DX} (empty)	N_{ii}	Adjustable parameters	Quality of the TDPPC Fit
(1) Simple donor	n	n	$n(0)$	No fit
(2) Simple acceptor (impossible; cannot achieve charge balance)	...	...	...	...
(3) Simple donor compensated by a net concentration of shallow acceptors	$n - N_{DX}^{+} + N_{SA}$	$n + 2N_{SA}$	$n(0)N_{SA}$, $N_{DX}^{+}(eq)$	No fit
(4) Simple acceptor which compensates a net concentration of shallow donors	$n - N_{DX}^{-} - N_{SD}$	$2N_{SD} - n$	$n(0)$, N_{SD}, $N_{DX}^{-}(eq)$	Poor fit
(5) Negative-U acceptor with intrinsic self-compensation (Refs. 10–11)	$n/2$	N_{Si}	$n(0)$	Good fit at short times
(6) Negative-U acceptor with intrinsic self-compensation (Refs. 10–11) plus a shallow donor	$(n - N_{SD})/2$	$N_{Si} + N_{SD}$	$n(0)$, N_{SD}	Excellent fit

be temperature-independent over the temperature range of interest.

IV. TDPPC SIMULATION OF EXPERIMENTAL DATA

For a quantitative simulation of the TDPPC data, Eqs. (1)–(8) must be solved simultaneously using the relations for N_{DX} (empty) and N_{ii} and the adjustable parameters given in Table I. Additionally, values for σ_∞ and E_{cap} must be specified. From Ref. 18, $\sigma_\infty = 5\times10^{-14}$ cm^2. E_{cap} is experimentally deduced from an Arrhenius plot of the slope of the linear portion of the TDPPC curve as shown in Fig. 2 where θ is given by

$$\theta = k_B \ln[(T)(\text{slope})]. \tag{9}$$

In Eq. (9), "slope" is the slope of the linear portion of the TDPPC curve and T in the logarithmic term arises from the temperature dependence of the thermal velocity and of the Brooks–Herring equation [Eq. (7)]. From the slope of this curve the capture barrier, $E_{cap} = 0.24$ eV, is found to be in good agreement with other workers.[18,19]

Returning to Table I, it is evident that simple donor or acceptor models for DX with or without compensation do not result in TDPPC simulations compatible with the experimental data, even when the model is formulated with three adjustable parameters.

In contrast, the model of Chadi and Chang[10,11] yields a *reasonable fit to the experimental data, at least over a range* of time, as shown in Fig. 3. According to this model,[10,11] persistent photoconductivity (PPC) arises according to the reaction

$$DX^{-} + d^{+} \underset{h\nu}{\rightarrow} 2d^{+} + 2e^{-} \tag{10}$$

where d^{+} denotes a normal, ionized shallow donor and DX^{-} represents the negatively charged DX state. TDPPC, according to this model, would involve the reverse of the reaction specified in Eq. (10). From Eq. (10) it is clear that two electrons are emitted to the conduction band for every

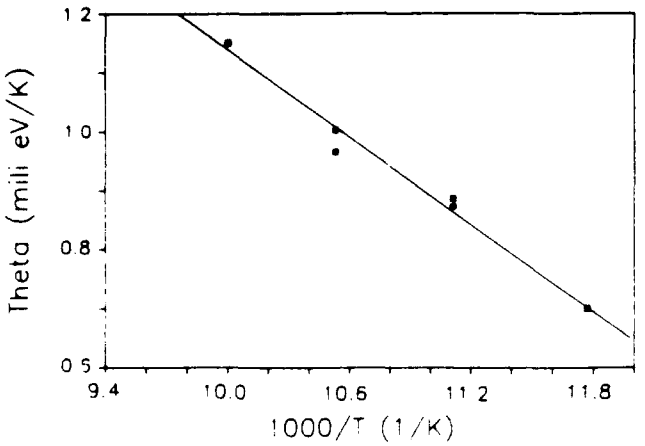


FIG. 2. Arrhenius plot of the slope of the inverse of the TDPPC which gives $E_{cap} = 0.24$ eV. Theta is related to the slope of the linear portion of the TDPPC curve.

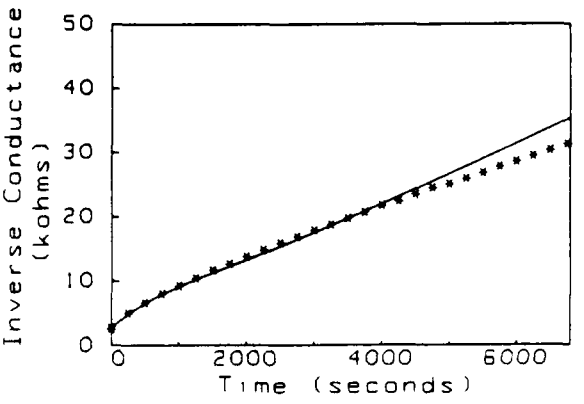


FIG. 3. A comparison of the simulation of Chadi and Chang's model (Refs. 10 and 11) (continuous line) to the experimental TDPPC data (asterisks) at a temperature of 85 K.

DX^- annihilated. Thus, one "empty" DX center exists for every two electrons such that $N_{DX}(\text{empty}) = n/2$, as indicated in Table I. It is also evident from Eq. (10) that the ionized impurity concentration is constant, independent of n, and equal to the total number of incorporated Si atoms, N_{Si}, as also indicated in Table I.

Although the fit shown in Fig. 3 based on Chadi and Chang's model is reasonable for short times, it is clear that it diverges significantly at longer times. Our computer simulation indicates that this divergence is due to more rapid electron recombination in the TDPPC simulation than that found experimentally. In order to obtain better agreement between the simulation and the experimental data, we postulate the presence of a shallow donor which remains ionized during the duration of the experiment. As shown in Fig. 4, such a postulate leads to excellent agreement between the simulation and the experimental data. The parameters used in the simulations shown in Fig. 4 are collected in Table II. Note that N_{SD} is equal to $1.3\times10^{16}/\text{cm}^3$ for all temperatures except for 100 K where the best fit occurs when $N_{SD} = 1.8\times10^{16}/\text{cm}^3$. Further TDPPC experiments over a wider range of temperature are required to establish why N_{SD} increases at 100 K, we suspect this may be associated with the neglect of the e_n^t term in Eq. (2).

V. EXISTENCE OF A SHALLOW DONOR

In order to obtain quantitative agreement over the full range of time between the TDPPC simulation and experimental data, we found it necessary to postulate the presence of a shallow donor. The purpose of this shallow trap is to account for the fact that the electron concentration, n, exhibits a greater degree of persistence than accounted for in the TDPPC simulation. This is illustrated in Fig. 5 where the simulated $n(t)$ curve is compared for Chadi and Chang's model with and without a shallow donor. It is obvious from Fig. 5 that the improvement in the TDPPC simulation through the inclusion of a shallow donor arises from the increased persistence of $n(t)$ at longer times. This increased persistence of $n(t)$ is also reflected in the increased mobility, $\mu(t)$, as shown in Fig. 6.

It should be recognized that our postulated shallow donor is photo-induced since this enhanced conductivity is only observed after photoexcitation (i.e., when the sample is cooled in the dark, the conductivity is unmeasurable). This inference of a photo-induced shallow donor is consistent with the photo-induced shallow electron trap deduced by Jia *et al.*[12] from transient capacitance measurements under strong illumination.

Several of the conclusions reached by Jia *et al.* are at odds, however, with our results. First, we deduce a photo-induced shallow donor concentration at approximately 1% of the doping level whereas Jia *et al.* concluded this concentration to be about 50%. Second, Jia *et al.* interpret data similar to Fig. 5 as arising from two simple exponentials whereas our formulation attributes the general shape of $n(t)$ to the nonlinear differential equation specified by Eqs. (2)–(4) and Table I. Third, Jia *et al.* assert that the long-time portion of the response is due to DX capture while the short-time response is associated with activated capture by the

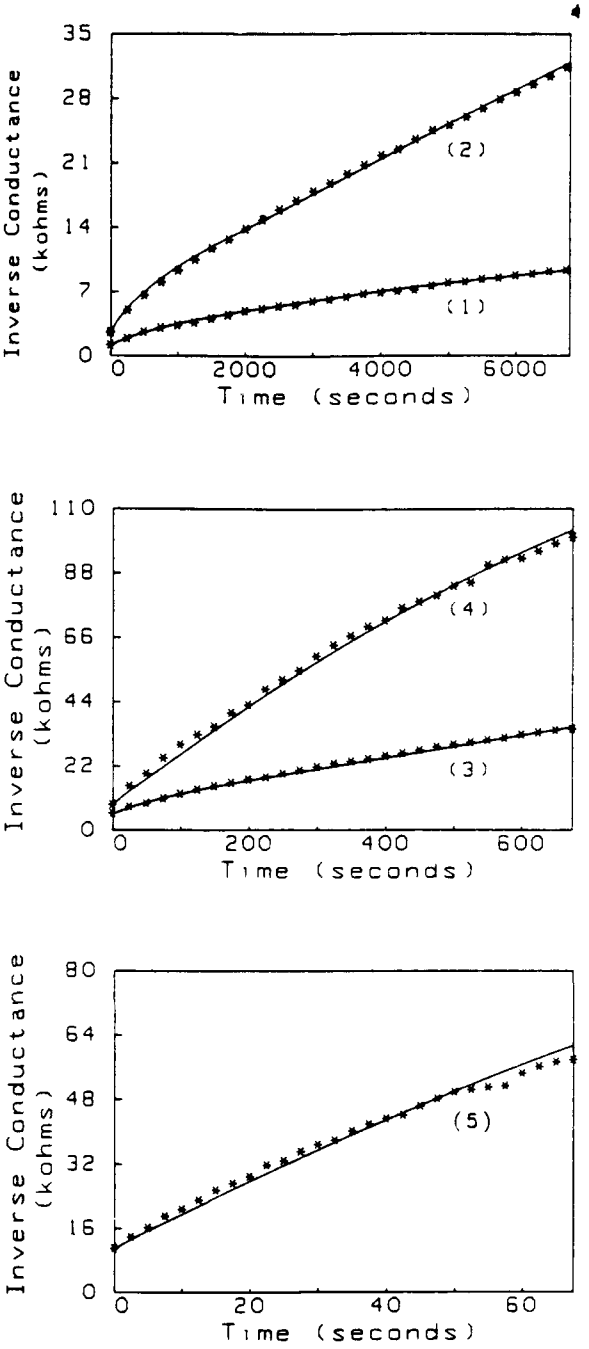


FIG. 4. A comparison of the simulation based on the model of Chadi and Chang (Refs. 10 and 11) plus a shallow donor (continuous line) to the experimental TDPPC data (asterisks) at temperatures of (1) 80 K, (2) 85 K, (3) 90 K, (4) 95 K, and (5) 100 K.

TABLE II. Parameters for TDPPC simulations based on the model of Chadi and Chang plus a shallow donor level.

Temperature (K)	$n(0)$ ($\times10^{17}$ cm^{-3})	N_{SD} ($\times10^{16}$ cm^{-3})	Figure
80	3.60	1.3	4(a)
85	2.42	1.3	4(b)
90	1.97	1.3	4(c)
95	1.33	1.3	4(d)
100	1.29	1.8	4(e)

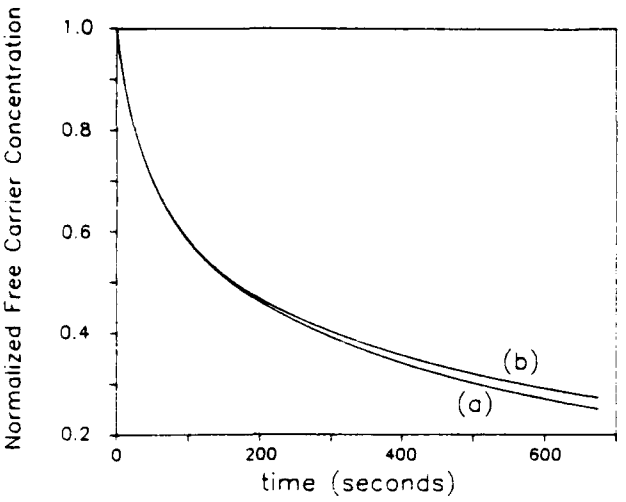


FIG. 5. Normalized transient decay of the electron concentration from a simulation based on Chadi and Chang's model (Refs. 10 and 11) without (a) and with (b) a shallow donor at a temperature of 90 K.

shallow donor. In contrast, we conclude from our TDPPC simulation that the short-time response is due to DX capture whereas the persistence of the long-time response is enhanced due to the absence of trapping of the shallow donor.

VI. CONCLUSIONS

Transient decay of persistent photoconductivity (TDPPC) measurements are analyzed in terms of a model in which the free carrier concentration decays by thermally activated capture of electrons into DX and a change in the mobility associated with a modification of the ionized impurity density concomitant with electron capture. Six models for DX are considered and simulations of the TDPPC data are performed for each model. Although qualitative agreement is obtained with Chadi and Chang's model,[10,11] quantitative agreement is only achieved by postulating the presence of a photo-induced shallow donor in addition to the DX model of Chadi and Chang. We believe that our results provide further support for the essential validity of the DX model of Chadi and Chang and that DX is acceptor-like, albeit a negative-U acceptor with intrinsic self-compensation. The existence of a photo-induced shallow donor has been previously reported by Jia *et al.*[12] The physical nature of this photo-induced shallow donor remains obscure at the present time.

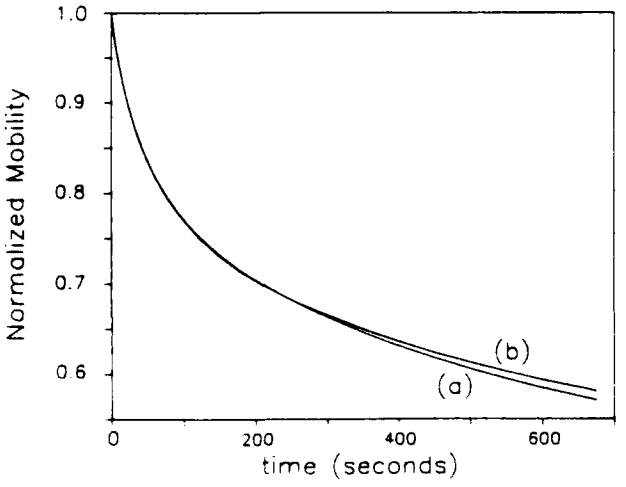


FIG. 6. Normalized transient decay of the mobility from a simulation based on Chadi and Chang's model (Refs. 10 and 11) without (a) and with (b) a shallow donor at a temperature of 90 K.

ACKNOWLEDGMENTS

We wish to thank Peter Neuchter for his many contributions to this work. Also, we thank Hyung-Mo Yoo, John Ebner, and John Arthur for MBE sample growth and Tom Plant for assistance with the optical setup. We also wish to thank Tom Theis for pointing out the importance of the Fermi-level correction to the capture cross section. This work was supported by the Air Force of Scientific Research under Contract Nos. AFOSR-86-0309 and AFOSR-89-0309. L. V. A. Scalvi was supported by CAPES (Brazil).

[1] R. J. Nelson, Appl. Phys. Lett. **31**, 351 (1977).
[2] D. V. Lang, R. A. Logan, and M. Jaros, Phys. Rev. B **19**, 1015 (1979).
[3] A. K. Saxena, Solid-State Electron. **25**, 127 (1982).
[4] D. M. Collins, D. E. Mars, B. Fischer, and C. Kocot, J. Appl. Phys. **54**, 857 (1983).
[5] H. Kunzel, A. Fischer, J. Knecht, and K. Ploog, Appl. Phys. A **32**, 69 (1983).
[6] N. Chand, T. Henderson, J. Klem, W. T. Masselink, R. Fischer, Y. C. Chang, and H. Morkoç, Phys. Rev. **30**, 4481 (1984).
[7] R. J. Nicholas, M. A. Brummell, J. C. Portal, G. Gregoris, S. Hersee, and J. P. Duchemin, Appl. Phys. Lett. **44**, 629 (1984).
[8] N. Chand, R. Fischer, J. Klem, T. Henderson, P. Pearah, W. T. Masselink, Y. C. Chang, and H. Morkoç, J. Vac. Sci. Technol. B **3**, 644 (1985).
[9] D. V. Lang in *Deep Centers in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, New York, 1986), p. 489.
[10] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
[11] D. J. Chadi and K. J. Chang, Phys. Rev. B **39**, 10063 (1989).
[12] Y. B. Jia, M. F. Li, J. Zhou, J. L. Gao, M. Y. Kong, P. Y. Yu, and K. T. Chan, J. Appl. Phys. **66**, 5632 (1989).
[13] H. C. Casey and M. B. Panish, *Heterostructure Lasers* (Academic, New York, 1978).
[14] T. M. Theis and P. M. Mooney, Mater. Res. Soc. Symp. Proc. (1989).
[15] T. W. Dobson and J. F. Wager, Mater. Res. Soc. Symp. Proc. (1989).
[16] H. J. Lee, L. Y. Juravel, J. C. Woolley, and A. J. Springthorps, Phys. Rev. B **21**, 659 (1980).
[17] R. A. Smith, *Semiconductors*, 2nd ed. (Cambridge, Cambridge, 1978).
[18] M. Takikawa and M. Ozeki, Jpn. J. Appl. Phys. **24**, 303 (1985).
[19] P. M. Mooney, N. S. Caswell, and S. L. Wright, J. Appl. Phys. **62**, 4786 (1987).

LOW-TEMPERATURE HOLE CAPTURE CROSS SECTION OF THE DX CENTER IN AlGaAs

S.B. Kim and J.F. Wager
Department of Electrical and Computer Engineering
Center for Advanced Materials Research
Oregon State University
Corvallis, OR 97331-3211

ABSTRACT

The minority carrier capture (MCC) technique is employed in order to estimate the low-temperature cross section of minority carrier holes into the DX center in Te- and Si-doped $Al_{0.3}Ga_{0.7}As$. The measured hole capture cross sections are approximately 2×10^{-19} cm^2 and 1.5×10^{-17} cm^2 for Te- and Si-doped $Al_{0.3}Ga_{0.7}As$, respectively, and are independent of temperature below approximately 70 K. The discrepancy in the magnitudes of these cross sections is explained in terms of multiphonon emission theory. The magnitude and temperature-dependence of these capture cross sections indicates that the rate-limiting step for hole capture involves DX in a neutral charge state.

Submitted to: Semiconductor Science and Technology

1. INTRODUCTION

For over a decade the physical nature of the DX center in AlGaAs has been a subject of great interest and controversy [1-7]. The DX center has attracted considerable attention because of its peculiar properties such as persistent photoconductivity (PPC) and a large Franck-Condon shift.

The purpose of this paper is to describe an experimental study in which the low-temperature hole capture cross section of the DX center is measured for Te- and Si-doped $Al_{0.3}Ga_{0.7}As$. The low-temperature hole capture cross section is evaluated using the minority carrier capture (MCC) technique [8]. The motivation for measuring the low-temperature capture cross section is to identify the charge state of the DX center which should aid in assessing its physical identity.

The DX hole capture cross section has been measured at intermediate temperatures (205-225 K) by Watanabe et al. [9] using deep-level transient spectroscopy (DLTS) with molecular beam epitaxy (MBE) $Al_{0.3}Ga_{0.7}As$ samples. In this range of temperature, the hole capture cross section was found to depend exponentially with temperature, consistent with capture by multiphonon emission (MPE) and to have an activation energy of 0.14 eV.

We have measured the low-temperature hole capture cross sections of the DX center for Te- and Si-doped $Al_{0.3}Ga_{0.7}As$ to be approximately $2x10^{-19}$ cm^2 and $1.5x10^{-17}$ cm^2, respectively. These capture cross sections are also found to be temperature-independent for less than about 70 K. According to the MPE theory of Ridley and Amato [10] and the experimental trends established by Henry and Lang [11] and by Hamilton et al. [8], low-temperature capture cross sections with magnitudes such as we measure and which are independent of temperature are consistent with capture by a neutral level. Thus, the primary conclusion of this study is that hole capture into the DX center occurs with DX in a neutral charge state.

Recently, the negative-U model for the DX center as proposed by Chadi and Chang [12,13] has gained wide acceptance as an explanation of DX behavior. Our conclusion that hole capture into the DX center involves DX in a neutral charge state can be reconciled with the model of Chadi and Chang if it is assumed that the MCC process is a two-step process in which the rate-limiting step is the capture of the second hole into a neutral state.

2. EXPERIMENTAL DETAILS

In the MCC process, photogenerated minority holes diffuse into the space-charge region of a Schottky barrier diode sample and are then captured by DX centers. Two types of starting material are used for the MCC experiments: (1) a 3 μm-thick, Te-doped, n-type (2.5×10^{17} cm^{-3}), $Al_{0.3}Ga_{0.7}As$ epitaxial layer is grown at Boeing by metalorganic chemical vapor deposition (MOCVD) on an n^+-GaAs substrate with a (100) orientation; and (2) a 2 μm-thick, Si-doped n-type (2×10^{17} cm^{-3}) is grown at Oregon State University by MBE using an n^+-GaAs substrate with a (100) orientation.

Au Schottky barrier contacts are evaporated onto the sample surface and are defined by either a shadow mask or by a lift-off process. Ohmic contact to the n^+ back side is accomplished using evaporated Ni/Au-Ge/Au metal layers and annealing 5 minutes at 420°C in forming gas. Processed wafers with Schottky barrier diodes are scribed into small pieces and mounted in a dual-in-line package.

The MCC experimental procedure essentially consists of measuring the capacitive rise times due to photexcitation with above- and below-bandgap light and also measuring the short-circuit current. The Schottky barrier sample is illuminated using a tungsten light source with a Jarrell-Ash monochrometer. The capacitance transient is monitored with a Hewlett-Packard 4280A fast capacitance meter and the short-circuit current is measured using a Hewlett-Packard 4140B picoammeter. Temperature control is achieved using an Air Products DF202 helium closed-cycle cryostat and a Scientific Instruments 5500 temperature controller. Equipment control and data acquisition is accomplished with a Hewlett-Packard 9326 desktop computer.

The experimental steps comprising the MCC experiment are as follows:

a) Cool the sample in the dark to a given temperature while applying a small forward bias to the sample.

b) Measure the capacitance transient under sub-bandgap illumination using a known photon flux.

c) Heat the sample to room temperature to reset it and then cool it to the same temperature as in a).

d) Measure the capacitance transient subject to bandedge illumination using the same photon flux as in b).

e) Measure the short-circuit current using bandedge light.

This procedure is then repeated for different temperatures.

3. MINORITY CARRIER CAPTURE: THEORY

The MCC expression for the hole capture cross section, σ_p, as derived by Hamilton et al. [8] is not relevant to the problem of minority hole capture by DX. Thus, a derivation of σ_p is provided using the same basic approach as employed by Hamilton et al.

The basic principle of the MCC technique is illustrated in Fig. 1. When a zero- or reverse-biased Schottky barrier is uniformly illuminated with bandedge light, a short-circuit photocurrent, composed of recombination and diffusion currents, is produced. Recombination current is generated when incident, band-edge photons create electron-hole pairs in the space-charge region which are subsequently separated by the internal space-charge field. Electron-hole pairs created within a diffusion length of the space-charge region edge can also contribute to the photocurrent if the photo-induced minority carriers diffuse into the space-charge region and then drift across the space-charge region to contribute to a diffusion current. In n-type semiconductors only holes (i.e., minority carriers) diffuse into the space-charge region whereas electrons are repelled by the potential barrier. Accordingly, the short circuit photocurrent is dominated by the diffusion current component as long as the minority carrier diffusion length is much larger than the width of the space-charge region, $L_p > W$.

As shown in Fig. 1, the charge state of the deep levels, which are initially occupied by electrons in the space-charge region, is changed by the capture of injected holes. The hole-capture is monitored by measuring the capacitance transient with a high-frequency capacitance meter since positive charge builds up in the space-charge region. Thus, the depletion region shrinks and the capacitance increases.

Excitation of the Schottky diode with weakly absorbing bandedge light can result in several excitation processes. These interactions are shown in Fig. 2, with corresponding transition rate parameters listed in Table 1. In Fig. 2, thermal and optical processes are indicated by straight and waved lines, respectively. The processes are described as follows:

Process 1: Optical absorption within a diffusion length of the space charge region creating an excess hole population inside the space charge region.

Process 2: Optical absorption in the space charge region creating electron-hole pairs separated by the field.

Process 3: Holes created by process 1 are captured by the electron-filled state (i.e., the most desirable outcome in MCC experimentation).

Process 4: Electrons created by process 2 are captured by empty (hole-filled) states.

Process 5: Direct electron photoionization.

Process 6: Direct hole photoionization.

Process 7: Thermal ionization of electrons.

Process 8: Thermal ionization of holes.

Note that processes 4, 6, and 8 tend to cause negative charge build-up in the depletion layer. Accordingly, the depletion width expands and the capacitance decreases. Conversely, processes 3, 5, and 7 tend to cause positive charge build-up in the depletion layer. Thus, the depletion width shrinks and the capacitance increases. In Table 1, standard notation is used, where ϕ and α denote, respectively, the photon flux and the absorption coefficient, c_n and c_p are, respectively, the electron and hole capture rates, e_n and e_p denote the corresponding emission rates, and the superscripts t and o denote thermal and optical processes, respectively. The incremental values Δn and Δp denote excess carrier densities produced by the photocurrent.

If the total concentration of the deep level is N_T, then $N_T = n_T + p_T$, where n_T and p_T are, respectively, the instantaneous concentration of trapped electrons and holes. If the illumination activates all the processes listed in Table 1, then the rate of change of the trapped electron density, $\frac{dn_T}{dt}$, will be

$\frac{dn_T}{dt}$ = (processes involving an increase in the trapped e^- concentration) - (processes involving a decrease in the trapped e^- concentration),

$\frac{dn_T}{dt}$ = (processes 8, 6, and 4) - (processes 7, 5, and 3)

and

$$\frac{dn_T}{dt} = \left(e_p^t + e_p^o + c_n\Delta n\right)\cdot p_T - \left(e_n^t + e_n^o + c_p\Delta p\right)\cdot n_T \qquad (1)$$

At temperatures below 80 K, the thermal ionization processes are negligible and $e_p^t = e_n^t \approx 0$. Note that in contrast to silicon and other binary compound semiconductors, process 6, hole photoionization, does not occur for AlGaAs even after DX is ionized because of large lattice relaxation. That is, the PPC effect cannot be optically quenched, as observed by Nelson [1]. Accordingly, process 6 can be ignored and $e_p^o \approx 0$. If the hole diffusion length, L_p, is much larger than the space-charge width, W, then the diffusion component of the photocurrent dominates and $\Delta p > \Delta n$. It is generally accepted that the electron capture cross section (σ_n) of DX centers at low temperature is extremely small (hence, PPC occurs). Thus, the hole capture rate (process 3) dominates over the electron capture rate (process 4) and $c_p\Delta p > c_n\Delta n$.

Given the above conditions, rate equation (1) can be rewritten as

$$\frac{dn_T}{dt} = -\left(e_n^o + c_p\Delta p\right)\cdot n_T \qquad (2)$$

which can be solved to obtain

$$\ln n_T = -\frac{t}{\tau_1} + c \qquad (3)$$

where

$$\tau_1 = \frac{1}{e_n^o + c_p\Delta p}\,.$$

The boundary condition of relevance is $n_T = N_T$ at $t = 0$; that is, the DX centers are filled with electrons. This condition can be achieved by application of a forward bias to the sample as it is cooled. Thus, $c = \ln N_T$, and taking the exponential of each side of (3) gives

$$n_T = N_T\, e^{-t/\tau_1} \tag{4}$$

also, $n_T = N_T - p_T$ so that

$$p_T(t) = N_T \left(1 - e^{-t/\tau_1}\right) . \tag{5}$$

The charge exchange kinetics are reflected directly in the high frequency capacitance per unit area of the barrier:

$$C(t) = \left(\frac{\epsilon\, \rho^+(t)}{2\,(V_{bi} \pm V)}\right)^{1/2} = B\left(\rho^+(t)\right)^{1/2} = Bq^{1/2}\left[N_{net}^+ + p_T(t)\right]^{1/2} , \tag{6}$$

and

$$B = \left(\frac{\epsilon}{2\,(V_{bi} \pm V)}\right)^{1/2} \quad \text{and} \quad \rho^+(t) = q\left[N_{net}^+ + p_T(t)\right] ,$$

where ρ^+ is the net positive space charge density in the space charge region, $V_{bi} \pm V$ is the total potential drop across the barrier, and N_{net}^+ is the net positive charge concentration in equilibrium. When $t = 0$, that is, before turning on the light, $C(t)$ is

$$C(0) = B\, q^{1/2}\left(N_{net}^+\right)^{1/2} \tag{7}$$

since $n_T = N_T$ at $t = 0$, and $p_T = 0$. Next, (6) and (7) may be manipulated [14] to yield

$$\frac{C^2(t) - C^2(0)}{C^2(0)} = \frac{\Delta C^2(t) + 2C(0)C(t) - 2C^2(0)}{C^2(0)} = \frac{p_T(t)}{N_{net}^+} \tag{8}$$

which leads to two limiting cases. First, if the capacitance change is much less than the initial capacitance, i.e., $\Delta C(t) \ll C(0)$, then the expression can be simplified as

$$\frac{C^2(t) - C^2(0)}{C^2(0)} \approx \frac{2\,\Delta C(t)}{C(0)} = \frac{p_T(t)}{N_{net}^+} . \tag{9}$$

Therefore, the capacitance change as a function of time is expressed as

$$\Delta C(t) = \frac{1}{2} C(0) \frac{p_T(t)}{N_{net}^+} . \tag{10}$$

Second, if the capacitance change is greater than the initial capacitance, $\Delta C(t) > C(0)$, then the expression is

$$\frac{C^2(t) - C^2(0)}{C^2(0)} = \frac{\Delta C^2(t) + 2\,\Delta C(t)C(0)}{C^2(0)} = \frac{p_T(t)}{N_{net}^+} \tag{11}$$

and the capacitance change as a function of time is expressed as

$$\Delta C^2(t) + 2\,\Delta C(t)C(0) = C^2(0) \frac{p_T(t)}{N_{net}^+} . \tag{12}$$

For the current investigation, the capacitance change is larger than the initial capacitance. Thus, from (5) and (12),

$$\Delta C^2(t) + 2\,\Delta C(t)C(0) = \frac{N_T}{N_{net}^+} C^2(0) \left(1 - e^{-t/\tau_{r1}}\right) \tag{13}$$

where $\tau_{r1} = (e_n^o + c_p \Delta p)^{-1}$ = capacitance rise-time measured under bandedge illumination (i.e., illumination at a photon energy slightly above bandgap where the absorption depth is large, ensuring that carrier generation occurs in the semiconductor bulk). The hole capture rate, c_p can be expressed as,

$$c_p = \upsilon_{th}(h^+) \cdot \sigma_p \tag{14}$$

so that using the definition of τ_{r1} and (14) and solving for the hole capture cross section gives,

$$\sigma_p(T) = \frac{\tau_{r1}^{-1} - \tau_{r2}^{-1}}{\upsilon_{th}(h^+) \cdot \Delta p} , \tag{15}$$

where τ_{r2}^{-1} is defined as e_n^o. τ_{r2} is the measured capacitance rise-time under sub-bandgap illumination. In (15), τ_{r1} and τ_{r2} are experimentally measurable values, while $\upsilon_{th}(h^+)$ is the hole thermal velocity expressed by

$$\upsilon_{th}(h^+) = \sqrt{\frac{3\ k_B T}{m_p^*}}\ , \tag{16}$$

where k_B = Boltzmann's constant and m_p^* = the effective mass of the hole.

It is then necessary to determine the injected hole concentration, Δp. If the hole diffusion length, L_p, is much larger than the depletion width, then the short-circuit current, J_{sc}, is considered to be composed of just diffusion current. Therefore,

$$J_{sc} = q\ D_p \left.\frac{dp}{dx}\right|_{x=0} \quad \text{where} \quad p = \Delta p\ e^{-x/L_p}\ , \tag{17}$$

and where x = 0 denotes the edge of the space charge region. Also,

$$\left|\frac{dp}{dx}\right|_{x=0} = \left|\Delta p \cdot \left(-\frac{1}{L_p}\right) \cdot e^{-x/L_p}\right|_{x=0} = \frac{\Delta p}{L_p}\ . \tag{18}$$

Substituting (18) into (17), expressing L_p in terms of D_p and the hole lifetime, τ_p, and solving for Δp gives

$$\Delta p = \frac{J_{sc}}{q\ (D_p/\tau_p)^{1/2}} \tag{19}$$

Finally, the hole capture cross section is obtained by combining (15) and (19), and is expressed as

$$\sigma_p(T) = \frac{q\ \left(\tau_{r1}^{-1} - \tau_{r2}^{-1}\right)\ (D_p/\tau_p)^{1/2}}{\upsilon_{th}(h^+)\ J_{sc}} \tag{20}$$

where

τ_{r1} = capacitance rise time under bandedge illumination,

τ_{r2} = capacitance rise time under sub-bandgap illumination,

τ_p = minority carrier lifetime,

$\upsilon_{th}(h^+)$ = thermal velocity of holes,

D_p = hole diffusivity, and

J_{sc} = short-circuit current density.

Equation (20) is the expression used to evaluate the low temperature hole capture cross section in the MCC experiment.

4. MINORITY CARRIER CAPTURE: RESULTS AND DISCUSSION

A. Evaluation of the Low-Temperature Capture Cross Section

To evaluate the low-temperature hole capture cross section using (20), the parameters τ_{r1}, τ_{r2}, τ_p, $\upsilon_{th}(h^+)$, D_p, and J_{sc} must be determined. τ_{r1}, τ_{r2}, and J_{sc} are obtained from experiment, while the remaining parameters are obtained as follows. The hole thermal velocity is given by (16) using m_p^* ($x = 0.3$) $\approx 0.666\, m_o$ and the relevant temperature as summarized in Tables 2 and 3. The hole diffusivity is estimated using the Einstein relation,

$$D_p = \mu_p\, k_B\, T/q\ , \tag{21}$$

and the hole mobility is assumed to be dominated by ionized impurity scattering so that μ_p is estimated using the Brooks-Herring equation [15],

$$\mu_p = \frac{64\sqrt{\pi}\ \epsilon^2\ (2k_BT)^{3/2}}{N_I\ q^3\ \sqrt{m_p^*}} \left[\ln\left\{ \frac{96\pi^2\ m_p^*\ k_B^2\ T^2\ \epsilon}{q^2\ h^2\ N_I} \right\} \right]^{-1}\ , \tag{22}$$

where ϵ, h, and q are the dielectric constant, Planck's constant, and the electronic charge, respectively, and N_I is the total concentration of ionized impurities. Finally, $\tau_p \approx 10$ nsec as reported by Ahrenkiel et al. [16].

The capacitance rise times, τ_{r1} and τ_{r2}, measured under bandedge and sub-bandgap illumination are obtained from capacitance transient measurements as shown in Fig. 3 for a Te-doped sample. The rise time is estimated from values of 10% to 90% of the capacitance transient.

All of the relevant parameters of (20), as well as the calculated capture cross sections for a variety of temperatures, are summarized in Tables 2 and 3 for Te- and Si-doped samples, respectively. As may be seen from these tables, the hole capture cross sections evaluated for the Te- and Si-doped samples are approximately 2.0×10^{-19} cm^2 and 1.5×10^{-17} cm^2, respectively, and are independent of temperature below ~ 70 K. Thus, while it can be concluded that both Te and Si DX centers have a small cross section for minority carrier hole capture, there is a discrepancy in the magnitudes of these capture cross sections.

B. MPE Theory Explanation of the Discrepancy of Measured Capture Cross Sections

This discrepancy in the magnitudes of the low-temperature capture cross sections for Te- and Si-doped samples can be explained in terms of nonradiative recombination by multiphonon emission (MPE) [17]. In general, the MPE capture cross section, σ, is expressed by [17]

$$\sigma = Af(0) , \tag{23}$$

where

A = a term involving only the electronic matrix elements of the transition, and

f(hv) = the optical lineshape for phonon-assisted absorption or emission transitions.

For MPE capture, the temperature-dependence of f(0) is given by [17]

$$f(0) = \frac{e^{-E_{cap}/k_B T^*}}{\sqrt{4\pi\, E_R k_B T^*}} \tag{24}$$

where E_R = relaxation energy, E_{cap} = capture barrier, and T^* = effective temperature. The effective temperature, T^*, is defined as [17]

$$k_B T^* = \frac{h\omega}{4\pi} \coth\left[\frac{h\omega}{4\pi\, k_B T}\right] . \tag{25}$$

At low temperature, the effective temperature, $k_B T^*$, is simplified as

$$k_B T^* = \frac{h\omega}{4\pi} , \tag{26}$$

since

$$\coth\left[\frac{h\omega}{4\pi\, k_B T}\right] \approx 1 \quad \text{with} \quad \frac{h\omega}{4\pi\, k_B T} > 1 . \tag{27}$$

Therefore, the MPE low-temperature capture cross section can be expressed as

$$\sigma_{LT} = \frac{A\theta^{-4\pi E_{cap}/h\omega}}{\sqrt{E_R\, h\omega}}, \qquad (28)$$

where $h\omega$ is an average phonon energy and the parameter, A, is an adjustable parameter which can vary up to one order of magnitude from the value of Henry and Lang [11] of 1.5×10^{-14} cm^2-eV which will be assumed in the subsequent analysis. To evaluate the theoretical low temperature capture cross section, the parameters E_{cap}, $h\omega$, and E_R must be known; these values are not known, however, for nonradiative hole capture by MPE.

As can be seen from an analysis of (28), σ_{LT} depends only weakly on E_R. Additionally, the average phonon energy, $h\omega$, should not depend strongly on the dopant. Thus, according to MPE theory, differences in σ_{LT} are most likely associated with small differences in E_{cap}. As an illustrative example, if we employ E_R = 0.75 eV, as used by Lang [17] in his MPE analysis of the DX center in Te-doped AlGaAs, $h\omega/2\pi$ = 40 meV, an average optical phonon energy for $Al_{0.3}Ga_{0.7}As$ [18], and E_{cap} = 0.14 eV, as measured by Watanabe et al. [9] for Si-doped $Al_{0.3}Ga_{0.7}As$, we obtain $\sigma_{LT} = 3\times10^{-17}$ cm^2, consistent with what we measure for Si-doped $Al_{0.3}Ga_{0.7}As$. If all of the parameters are held constant except for the capture barrier which is changed to E_{cap} = 0.25 eV, we obtain $\sigma_{LT} = 1\times10^{-19}$ cm^2, consistent with our measured value for Te-doped $Al_{0.3}Ga_{0.7}As$. Thus, a two order of magnitude change in σ_{LT} can be accounted for by 110 meV change in E_{cap}. Therefore, we conclude that the differing experimental values of the low-temperature capture cross sections for Te- and Si-doped samples are consistent with MPE theory. In particular, these differences are likely due to small differences in the capture barrier, E_{cap}, with Te-derived DX exhibiting a larger barrier to hole capture.

C. The Charge State of the DX Center

Low-temperature hole capture cross sections measured by the MCC technique for Te- and Si-doped samples are shown in Fig. 4 and are compared to intermediate-temperature hole capture cross sections measured by DLTS [9]. It should be noted that the data of Watanabe et al. shown in Fig. 4 imply a smaller low-temperature hole capture cross section for Si-doped $Al_{0.3}Ga_{0.7}As$ than what we deduce from our MCC

experiment. In contrast, Brunthaler et al. [19] estimate σ_{LT} (30K) $\geq 2x10^{-16}$ cm^2 for Si-doped $Al_{0.34}Ga_{0.66}As$ from time-resolved photoluminescence studies. We attribute this lack of agreement in the estimation of σ_{LT} between various researchers to differences in samples and in the experimental techniques employed. We also note that the temperature-independence of σ_{LT} is more important for the assessment of the nature of DX center than these relatively minor differences in the estimated magnitude of σ_{LT}.

For comparative purposes, trends in the capture rate as a function of temperature according to the MPE theory as proposed by Ridley and Amato [10] are shown in Fig. 5. According to this theory, at low temperature the capture cross section is determined solely by the charge of the deep level if capture occurs by MPE. The capture cross section increases as temperature decreases for Coulombically attractive capture. In contrast, neutral capture is temperature-independent whereas repulsive capture yields a decreasing cross section with decreasing temperature. In order to facilitate a comparison of our experimental data to trends predicted by Ridley and Amato, we include the hole capture rate, c_p [i.e. $c_p = \upsilon_{th}(h^+)\sigma_p$], in Tables 2 and 3. Note from Tables 2 and 3 that, to within experimental accuracy, the hole capture rate is temperature-independent from 30-70K for both Te- and Si-doped $Al_{0.3}Ga_{0.7}As$; this is consistent with neutral capture, according to the MPE theory of Ridley and Amato.

Additionally, Henry and Lang [11] and Hamilton et al. [8] show experimentally for 15 defects in GaAs and GaP that the low temperature capture cross section for Coulombically attractive centers is large ($\sim 10^{-13}$ - 10^{-15} cm^2) and increases with decreasing temperature while it is small ($\sim 10^{-18}$ - 10^{-21} cm^2) for neutral capture and is temperature-independent at low temperature as shown in Table 4.

In summary, MCC experiments indicate the low temperature minority capture cross section to be approximately $2.0x10^{-19}$ and $1.5x10^{-17}$ cm^2 for Te- and Si-doped samples, respectively, and to be temperature-independent. The MPE theory of Ridley and Amato [10] and the experimental trends of Henry and Lang [11] and of Hamilton et al. [8] indicate that because the MCC capture cross sections are of the magnitude measured and temperature-independent, DX capture must involve a neutral charge state.

This conclusion, that DX hole capture involves a neutral charge state, appears to be in conflict with the DX model of Chadi and Chang [12,13]. According to the model of Chadi and Chang, the MCC process can be described by:

$$2\,h^{+} + DX^{-} = d^{+}\,. \tag{29}$$

This is the reverse of the DX formation process which is described by:

$$d^{+} + 2\,e^{-} = DX^{-}\,. \tag{30}$$

Thus the MCC process would appear to involve Coulombically attractive capture in terms of the model of Chadi and Chang. However, the MCC experimental results can be reconciled with the model of Chadi and Chang if it is postulated that the MCC process can be described by a two-step process:

$$\begin{aligned} &\text{step 1 :}\quad h^{+} + DX^{-} = DX^{\circ} \\ &\text{step 2 :}\quad h^{+} + DX^{\circ} = d^{+}\ \text{(rate-limiting)} \end{aligned} \tag{31}$$

where DX° denotes an intermediate, unstable state of the DX center which exists for only a brief duration during the transition from DX^{-} to d^{+}. Step 2 is assumed to be the slower or rate-limiting process which dominates the overall MCC process. Thus, the results of the current investigation are compatible with the model of Chadi and Chang [12,13] for DX if the rate-limiting step for hole capture is identified as due to capture into the neutral DX° state.

5. CONCLUSIONS

An experimental study is presented in which the low-temperature capture cross sections for Te- and Si-doped $Al_{0.3}Ga_{0.7}As$ are estimated by the MCC technique to be approximately 2×10^{-19} cm^2 and 1.5×10^{-17}, respectively, and to be independent of temperature below about 70 K. In light of the MPE theory of Ridley and Amato [10] and the experimental trends established by Henry and Lang [11] and Hamilton et al. [8], we conclude from the magnitude and temperature-dependence of the measured low-temperature capture cross section that hole capture involves the DX center in a neutral charge state. This conclusion is compatible with the DX model of Chadi and Chang [12,13] if it can be assumed that hole capture occurs by a two-step sequential process in which the second step involves the neutral DX° state and that this is the rate-limiting step of the hole capture process.

ACKNOWLEDGMENTS

The authors wish to thank V. Sundaram of the Boeing Electronics Company for providing the MOCVD samples, L. Ungier and J.R. Arthur for MBE samples, and T.K. Plant for assistance and advice with the optical set up. This work was supported by the Air Force Office of Scientific Research under contract AFOSR 89-0309.

REFERENCES

[1] Nelson R J 1977 *Appl. Phys. Lett. 31* 351

[2] Lang D V, Logan R A, and Jaros M 1979 *Phys. Rev. B 19* 1015

[3] Lang D V 1986 in *Deep Centers in Semiconductors* ed by S T Pantelides (New York:Gordon and Breach) p 489

[4] Theis T N 1987 *Inst. Phys. Conf. Ser. No. 91* 1

[5] Bhattacharya P 1988 *Semicond. Sci. Technol. 3* 1145

[6] 1990 *Physics of DX Centers in GaAs Alloys*, ed by J C Bourgoin *Solid State Phenomena 10*

[7] Mooney P M 1990 *J. Appl. Phys. 67* R1

[8] Hamilton B, Peaker A R, and Wight D R 1979 *J. Appl. Phys. 10* 6373

[9] Watanabe M O, Ahizawa Y, Sugiyama N, and Nakanisi T 1986 *Inst. Phys. Conf. Ser. No. 83* 105

[10] Ridley B K and Amato M A 1981 *J. Phys. C. 14* 1255

[11] Henry C H and Lang D V 1977 *Phys. Rev. B 15* 989

[12] Chadi D J and Chang K J 1988 *Phys. Rev. Lett. 61* 873

[13] Chadi D J and Chang K J 1989 *Phys. Rev. B 39* 10063

[14] Kim S B 1991 *Ph D Thesis* Oregon State University

[15] Smith R A 1978 *Semiconductors* (Cambridge:Cambridge University Press) p 253

[16] Ahrenkiel R K, Dunlavy D J, Loo R Y, and Kamath G S 1988 *J. Appl. Phys. 63* 5174

[17] Lang D V 1980 *J. Phys. Soc. Japan, Suppl. A 49* 215

[18] Jusserand B and Sapriel J 1981 *Phys. Rev. B 12* 7194

[19] Brunthaler G and Ploog K 1989 *Phys. Rev. Lett. 63* 2276

[20] Lang D V and Logan R A 1975 *J. Electron. Mater. 4* 1053

[21] Dean P J, Skolnick M S, Uihlein C., and Herbert D C 1983 *J. Phys. C. 16* 2017

Table 1. Notation for transition probabilities corresponding to the optical interactions shown in Fig 2; after Hamilton et al. (1979)

Transition	Process	Transition Probability per unit Time
Optical absorption	1 & 2	Proportional to $\phi\alpha$
Hole capture	3	$C_p = \sigma_p V_{th}(h^+)\Delta p$
Electron capture	4	$C_n = \sigma_n V_{th}(e^-)\Delta n$
Electron photoionization	5	$e_n^o = \phi \sigma_n^o$
Hole photoionization	6	$e_p^o = \phi \sigma_p^o$
Electron thermal ionization	7	$e_n^t = \sigma_n V_{th} N_c \exp\left(\frac{-(E_c - E_T)}{k_B T}\right)$
Hole thermal ionization	8	$e_p^t = \sigma_p V_{th} N_v \exp\left(\frac{-(E_T - E_v)}{k_B T}\right)$

Table 2. MCC experimental results for Te-doped $Al_{0.3}Ga_{0.7}As$

Parameters	Temperature			
	30 K	50 K	60 K	70 K
τ_{r1} (sec)	95.7	99	105.6	108.9
τ_{r2} (sec)	124.3	125.4	128.7	132
J_{sc} (A/cm^2)	$1.3x10^{-6}$	$1.3x10^{-6}$	$1.3x10^{-6}$	$1.3x10^{-6}$
$\upsilon_{th}(h^+)$ (cm/sec)	$4.5x10^{6}$	$5.8x10^{6}$	$6.4x10^{6}$	$6.9x10^{6}$
D_p (V^2/cm-sec)	0.102	0.34	0.52	0.75
μ_p (V/cm-sec)	39.5	78.51	100.5	123.9
$\sigma_p(T)$ (cm^2)	$2.10x10^{-19}$	$2.63x10^{-19}$	$2.36x10^{-19}$	$2.48x10^{-19}$
$c_p(T)$ (cm^3/sec)	$9.5x10^{-13}$	$1.5x10^{-12}$	$1.5x10^{-12}$	$1.7x10^{-12}$

Table 3. MCC experimental results for Si-doped $Al_{0.3}Ga_{0.7}As$

Parameters	Temperature				
	40 K	50 K	60 K	70 K	80 K
τ_{r1} (sec)	20.0	22.7	23.6	20.8	19.0
τ_{r2} (sec)	> 150	> 150	> 150	> 150	> 150
J_{sc} (A/cm^2)	0.35	0.27	0.37	0.45	0.33
$\upsilon_{th}(h^+)$ (cm/sec)	$5.2x10^{6}$	$5.8x10^{6}$	$6.4x10^{6}$	$6.9x10^{6}$	$7.4x10^{6}$
D_p (V^2/cm-sec)	0.1526	0.2596	0.4012	0.5797	0.7976
μ_p (V/cm-sec)	44.2	60.2	77.5	96.0	115.6
$\sigma_p(T)$ (cm^2)	$1.3x10^{-17}$	$1.66x10^{-17}$	$1.3x10^{-17}$	$1.4x10^{-17}$	$2.4x10^{-17}$
$c_p(T)$ (cm^3/sec)	$6.8x10^{-11}$	$9.6x10^{-11}$	$8.3x10^{-11}$	$9.7x10^{-11}$	$1.8x10^{-10}$

Table 4. A summary of experimentally determined low-temperature capture-cross section trends for GaAs and GaP

Semiconductor	Defect	Type of Capture	σ_{LT} (cm^2)	σ_{LT}-Trend with Decreasing Temperature[a]	Nature of Coulombic Interaction	References
GaAs	A	hole	$> 5 \times 10^{-15}$	increasing (220 K)	attractive	11, 20
	A	electron	$\sim 5 \times 10^{-18}$	constant	unknown	11
	Cr	electron	$< 10^{-19}$	decreasing (250 K)	neutral	11, 20
	Fe	electron	$< 10^{-19}$	decreasing (250 K)	neutral	11, 20
	B	electron	$\sim 10^{-21}$	constant	unknown	11
	B	hole	$> 2 \times 10^{-15}$	increasing (260 K)	attractive	11, 20
	E3	electron	$< 5 \times 10^{-17}$	decreasing (210 K)	unknown	11
	Cu	hole	$> 2 \times 10^{-15}$	increasing	attractive	11, 20
	O, EL2	electron	$< 5 \times 10^{-18}$	decreasing (110 K)	neutral	11
GaP	O	hole (σ_{p1})	4×10^{-21}	constant	neutral	11, 21
	O	hole (σ_{p2})	$> 10^{-14}$	increasing (200 K)	attractive	11, 21
	O	electron (σ_{n2})	2×10^{-20}	constant	neutral	11, 21
	ZnO	electron	$< 10^{-15}$	decreasing (180 K)	neutral[b]	11
	0.95 eV hole trap	hole	$> 10^{-14}$	increasing (80 K)	attractive	8
	0.75 eV hole trap	hole	$> 10^{-12}$	increasing (120 K)	attractive	8

Note:

(a): Increasing, decreasing, and constant refers to the trend in the experimental capture cross section at the lowest measurement temperature. The lowest measurement temperature is indicated in parentheses for traps which do not show a constant capture cross section.

(b): ZnO is a two-defect, isoelectronic trap in GaP. The Coulombic interaction during electron capture is more complicated than for a simple trap because of the dipole moment of the ZnO isoelectronic trap.

FIGURE CAPTIONS

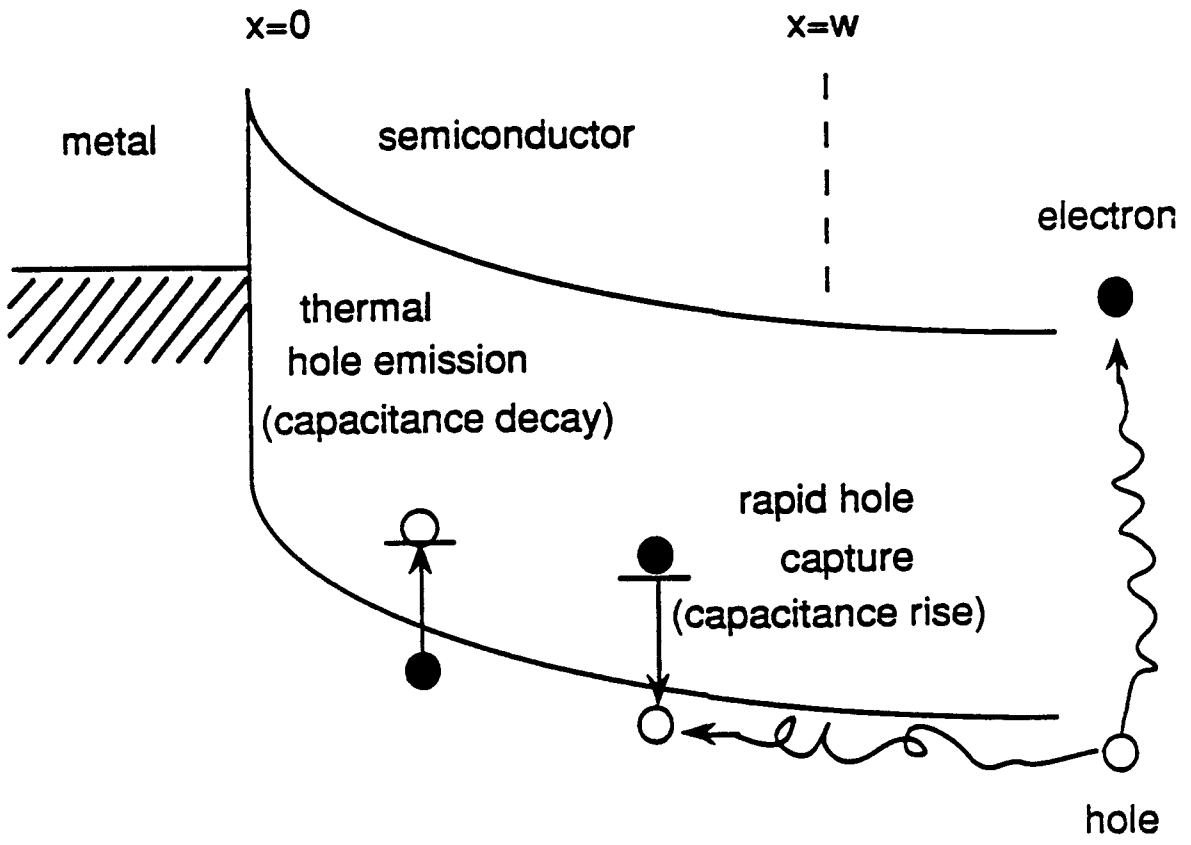
Figure 1. Basic principle of minority carrier capture; after Hamilton et al. (1979).

Figure 2. Optical interactions resulting from near bandgap illumination; after Hamilton et al. (1979).

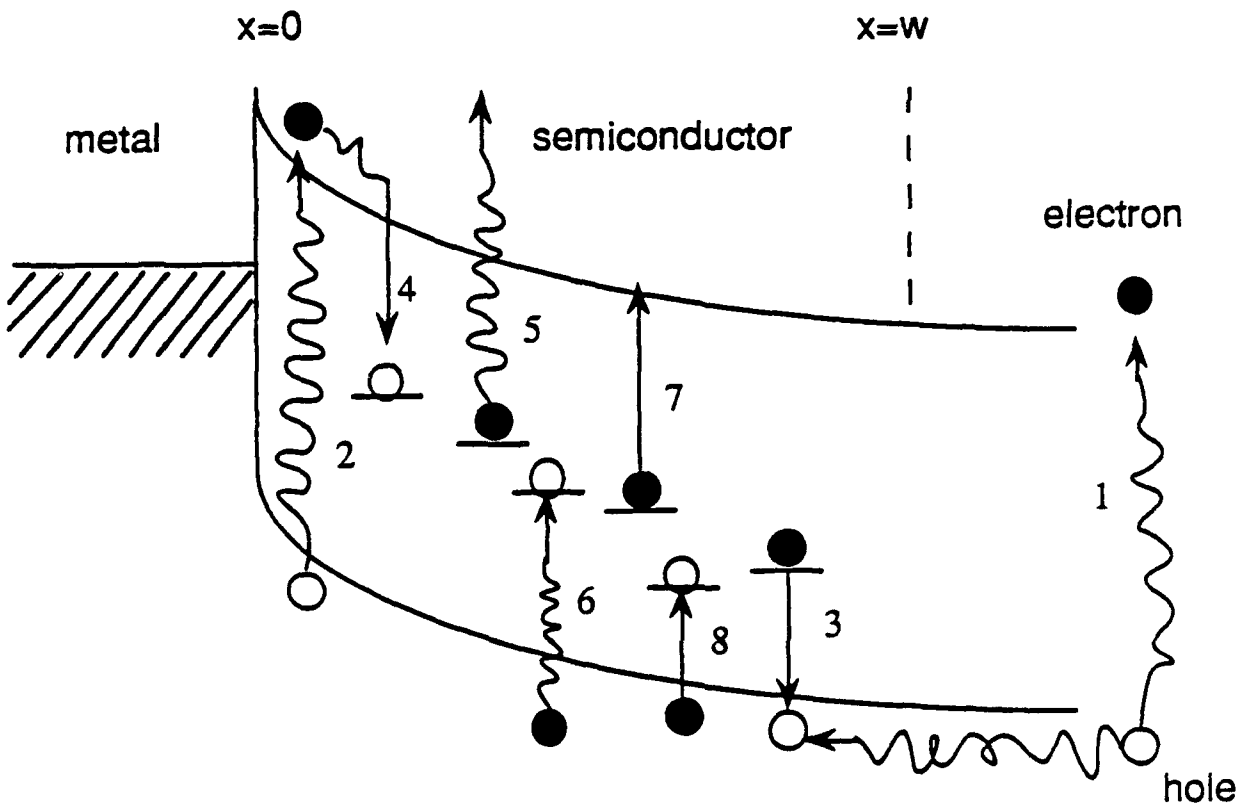
Figure 3. Capacitance rise time measured at 50 K for a Te-doped sample: (a) under near-bandgap illumination, $h\nu$ = 1.97 eV, and (b) sub-bandgap illumination, $h\nu$ = 1.38 eV. The energy bandgap of the n-$Al_{0.3}Ga_{0.7}As$ at 50 K is 1.94 eV.

Figure 4. Low-temperature hole capture cross sections obtained from MCC experiments for Te- and Si-doped samples. The data designated by +'s is from Watanabe et al. (1986), in which DLTS was employed to measure the intermediate-temperature hole capture cross section.

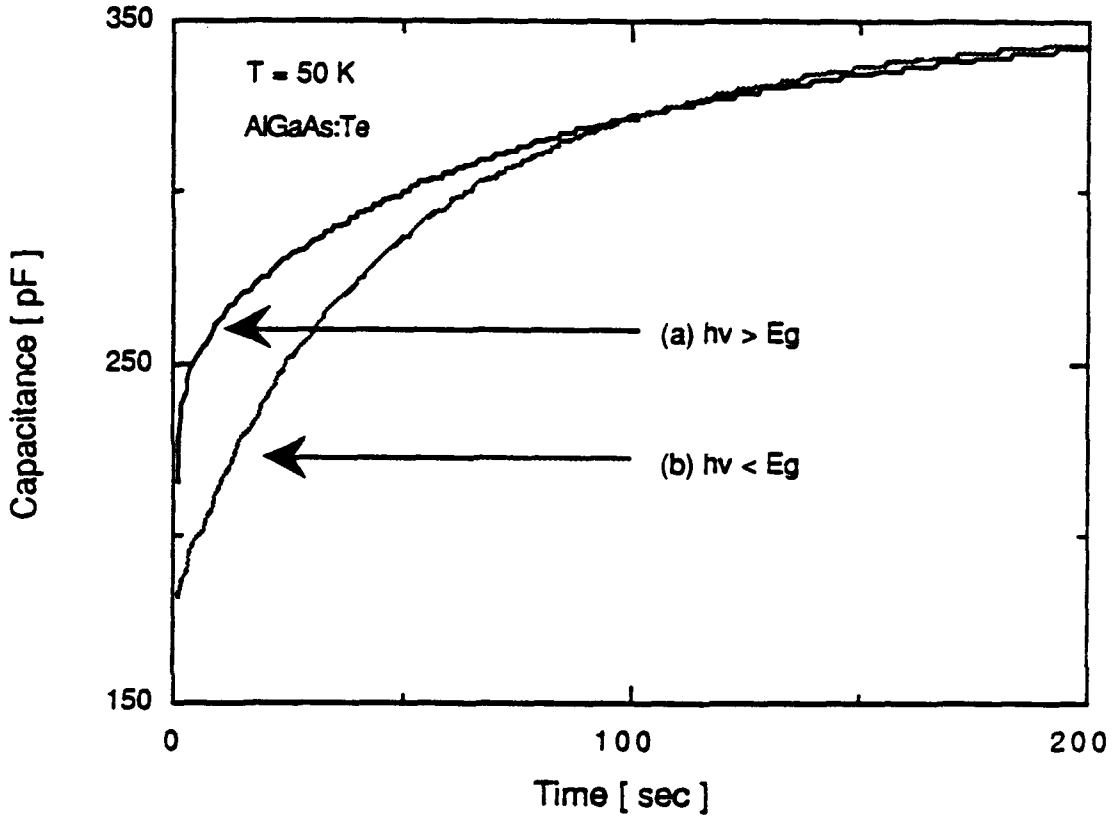
Figure 5. Capture rate as a function of the charge state of the deep level over a wide temperature range; after Ridley and Amato (1981).



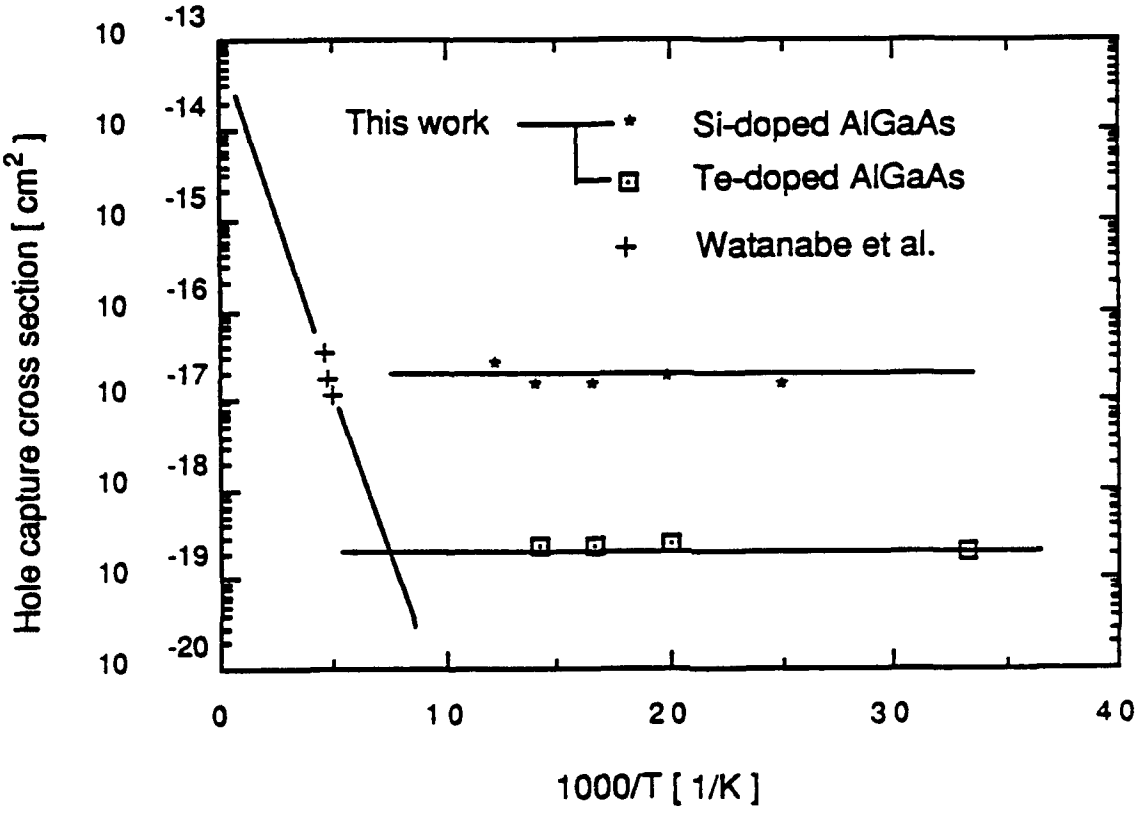
Kim & Wager Figure 1



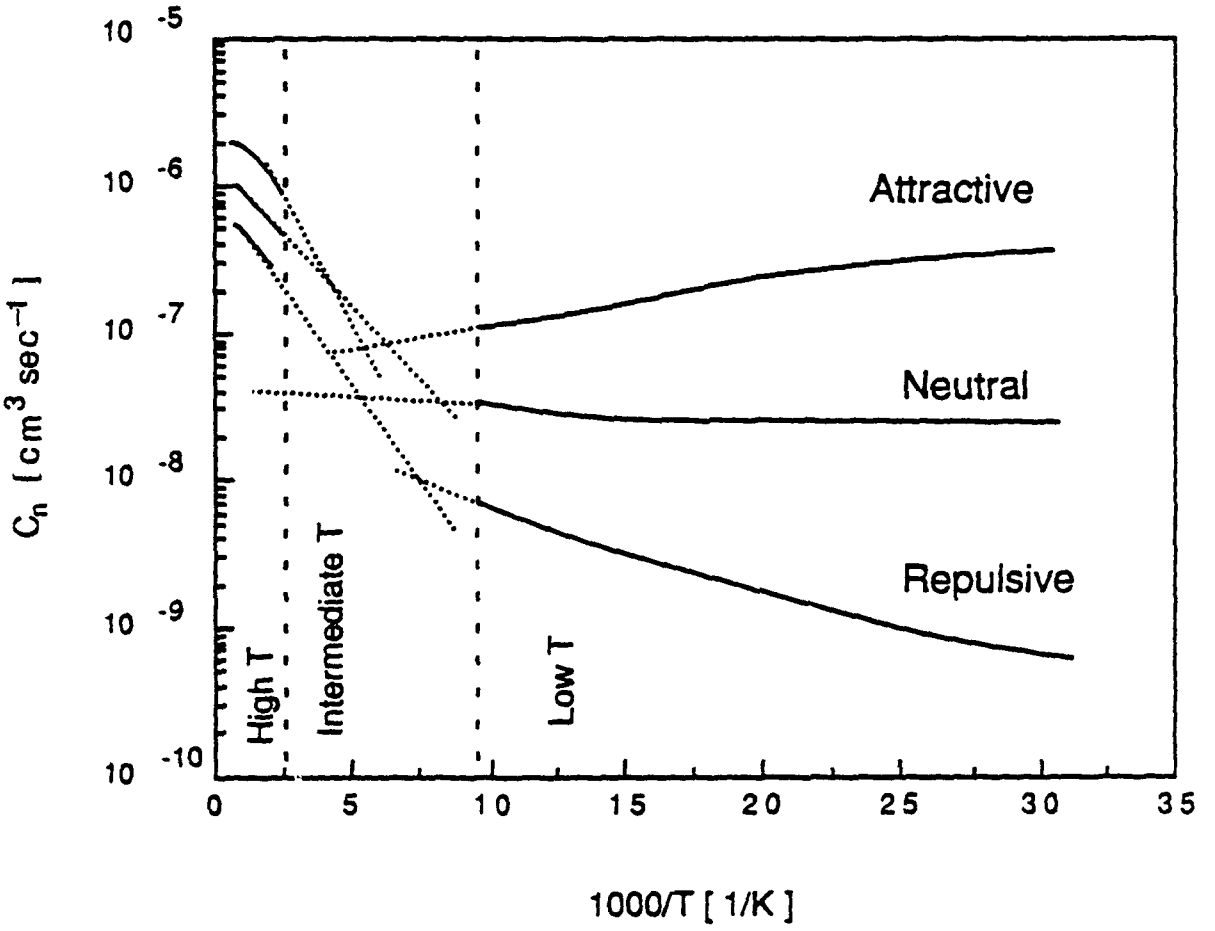
KIM AND WAGER FIGURE 2



KIM & WAGNER FIGURE 3



Kim & Wagner Figure 4



Kim & Wagner Figure 5

Entropy of migration for atomic hopping

T. W. Dobson,* J. F. Wager, and J. A. Van Vechten
Center for Advanced Materials Research, Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331-3202

(Received 13 June 1988; revised manuscript received 6 April 1989)

The entropy of migration of an atom undergoing an atomic hop into a nearest-neighbor vacancy is calculated under the assumption that the entropy is translational at the saddle-point configuration and vibrational at the initial, equilibrium state prior to the hop. These contributions to the entropy of migration are calculated using standard statistical thermodynamic expressions, assuming that the translational entropy is given by the particle-in-a-box approximation and the vibrational entropy is that of a simple harmonic oscillator vibrating at the Debye frequency. This formulation of the entropy of migration is then quantitatively applied to explain the abnormal prefactors experimentally deduced in InP drain-current–drift measurements, in deep-level-defect–transformation kinetic studies of the metastable M center in InP, and in Si and Ge self-diffusion experiments.

I. INTRODUCTION

The kinetics of solid-state processes such as diffusion and defect-transformation reactions are often found to exhibit the Arrhenius form,[1]

$$k = A \exp(-E_a/k_B T) , \quad (1)$$

where k is the rate constant, A is the prefactor, E_a is the activation energy, and k_B is Boltzmann's constant. Adopting the thermodynamic formulation of rates,[1] it is possible to make the following identifications:

$$A = A_0 \exp(\Delta S_a/k_B) , \quad (2)$$

$$E_a = \Delta H_a , \quad (3)$$

where A_0 can loosely be interpreted as an attempt frequency and ΔS_a and ΔH_a are, respectively, the entropy and enthalpy of activation. These activation quantities represent the difference of the reactant at its saddle-point configuration with respect to its initial configuration.

The prefactor A has often been observed experimentally to be abnormally large, or alternatively, abnormally small, in a variety of solid-state processes such as self-diffusion[2–7] and deep-level–defect transformations.[8–14] The unusual magnitude of the prefactor is attributed to the correspondingly abnormal magnitude of the entropy of activation. The physical reason for the unusual magnitudes of these empirically determined activation entropies (i.e., $\Delta S_a/k_B > 5$ or $\Delta S_a/k_B < -5$) is generally controversial or unexplained in the prior literature and the subject of this work.

Three contributions to the entropy of activation have been discussed in the literature: (1) A vibrational contribution due to a modification of the vibrational frequencies of the lattice when the reactant achieves its saddle-point configuration,[15,16] (2) a configurational entropy due to the different multiplicity of bonding configurations (e.g., Jahn-Teller distortion) at the saddle point compared to the initial state,[17] and (3) the entropy of ionization due to the softening of the lattice when an electron or hole is emitted from a defect state compared to when it is localized at that state.[16,18]

Note that ionization must occur concomitant with the process of interest for mechanism 3 to appear in the corresponding activation barrier. In the prior literature, the first two mechanisms are generally regarded as producing a $\Delta S_a/k_B < 5$ and the third mechanism is usually either neglected or argued not to contribute on the ground that ionization is not correlated with the process. It has often been said a value of $\Delta S_a/k_B > 5$ for a simple point-defect migration process is "unphysical;" see, for example, Refs. 4 and 5.

The purpose of the work discussed herein is to propose a physical mechanism associated with atomic migration which is capable to explaining the very large magnitudes of the activation entropies $\Delta S_a/k_B \gg 5$. The proposal is motivated by our own observation of $\Delta S_a/k_B \sim 15$ for certain defect-transformation processes in InP.

Our explanation is based upon the ballistic-model (BM)[16,19,20] hypothesis that atomic migration into a vacancy on a nearest-neighbor site may be calculated by assuming that the migration energy is kinetic in origin. Within the context of this model, the migrating atom must attain a sufficient kinetic energy to successfully transverse the saddle-point configuration. In the spirit of the BM, we propose that the activation entropy of a process involving atomic migration (e.g., diffusion, metastable defect transformations) will be determined (at least to a large extent) by the difference of the translational entropy of the hopping atom in its saddle-point configuration and the vibrational entropy of the hopping atom in its equilibrium position prior to the hop. These two contributions to the activation entropy are calculated using standard statistical thermodynamic expressions.

Using this formulation for the entropy of activation of atomic migration, we provide quantitative justifications for experimentally obtained prefactors for the activation entropy for phosphorus-vacancy nearest-neighbor hopping contributions to drain-current drift in InP metal-

insulator-semiconductor (MIS) capacitors and field-effect transistors,[13,14] for the prefactors found in defect-transformation experiments involving the M center[8-12] in InP, and for the self-diffusion[4-7] of Si and Ge.

II. ENTROPY OF MIGRATION FOR NEAREST-NEIGHBOR HOPPING

The physical picture adopted in the BM is as follows. At temperatures above the Debye temperature θ_D, the thermal vibrations of the atoms surrounding a vacancy fluctuate with a characteristic frequency given approximately by the Debye frequency $\nu_D = k_B\theta_D/h$, where h is Planck's constant. Periodically these thermal fluctuations conspire to provide a path for an atom to hop into the vacancy at a cost of very little potential energy. The venue for this path is limited by the thermal motion of the surrounding atom to a period of order the zone-boundary phonon period of the host lattice, which is proportional to the Debye period ν_D^{-1}. For an atom to successfully migrate it must make the hop in a time less than the lifetime of the favorable venue. This requires a minimum velocity of the hopping atom to be

$$v = \nu_D d \ , \tag{4}$$

where d is the distance between lattice sites. This velocity corresponds to a kinetic energy

$$E_{\rm kin} = \tfrac{1}{2} m v^2 \ , \tag{5}$$

where m is the mass of the hopping atom. Thus, since the potential energy that the atom must overcome is assumed to be quite small during this favorable venue, the enthalpy of migration is given by

$$\Delta H_m = \tfrac{1}{2} m v^2 = \tfrac{1}{2} m (F d \nu_D)^2 \tag{6}$$

where F is a geometric constant equal to 0.9 in the case of a diamond or zinc-blende lattice.[19]

In order to calculate the entropy of migration, we extend the BM treatment. At temperatures above the Debye temperature an atom on a normal lattice site can be modeled as a simple harmonic oscillator vibrating at the Debye frequency. The entropy associated with this vibrating atom can be calculated from statistical thermodynamics and is given by[21]

$$\frac{S_v}{k_B} = 3\left\{ \frac{h\nu_D/k_BT}{\exp(h\nu_D/k_BT)-1} - \ln\left[1-\exp\left[\frac{-h\nu_D}{k_BT}\right]\right]\right\} , \tag{7}$$

where the factor 3 is due to the three degrees of freedom of the oscillating atom.

According to the BM, during the limited venue that the atomic migration actually occurs, i.e., when an atom moves through the saddle-point configuration, the hopping atom is essentially a free particle moving ballistically. Thus, for this brief period of time the vibrational modes are replaced by a translational mode. We can calculate the translational entropy using the particle-in-a-box approximation,[21]

$$\frac{S_t}{k_B} = \ln\left[\frac{(2\pi e m k_B T)^{3/2}}{h^3} V\right] , \tag{8}$$

where $e = 2.718$ and V is the volume of the box which we take to be the volume occupied by two nearest-neighbor atoms. In the present work we estimate the volume of the box as the atomic volume of two atoms as defined by the Van Vechten–Phillips tetrahedral radii.[22] There are other equally reasonable ways of approximating the box volume, such as a narrow channel geometry in which the hopping atom is restricted in directions perpendicular to the migration path. We note that either method of calculation yields approximately the same estimate of the box volume. Furthermore, it is clear from Eq. (8) that the entropy depends weakly (i.e., logarithmically) on the box volume.

The entropy of migration of an atom hopping into a vacancy is thus given by,

$$\Delta S_m = S_t - S_v \ . \tag{9}$$

Therefore, our formulation for the entropy of atomic migration is given by Eqs. (7)–(9). We will now employ these equations to explain the abnormal activation entropies found experimentally in defect-transformation reactions and self-diffusion.

We note that our formalism is implicitly based on a constant volume ensemble, whereas the experimental data we wish to consider are obtained under constant pressure. Thus, it is necessary to distinguish between constant volume and constant pressure entropies.[23] We do not correct for this difference since these corrections are rather small, less than k_B, for the semiconductors of interest in this work.

We also note that our formulation for the migrational entropy differs from that normally employed,[15,23]

$$\frac{\Delta S_m}{k_B} = \sum_i \ln\left[\frac{\omega_i}{\omega_i'}\right] , \tag{10}$$

where ω_i and ω_i' are the normal-mode frequencies on the initial and saddle-point configurations, respectively. Equation (10) is valid only in the context of the harmonic or quasiharmonic approximation and is deduced assuming that the migrational barrier is best calculated assuming it is potential energy in origin. Thus, the essential difference between our approach and the conventional approach is whether the migrational barrier is more readily calculated as a potential or kinetic energy barrier. It is difficult to believe that modifications of the vibrational frequencies of atoms surrounding a vacancy during atomic hopping could be of a sufficient magnitude to account for migrational entropies as large as $15k_B$.

III. ENTROPY OF MIGRATION EXAMPLES

A. Phosphorus-vacancy nearest-neighbor hopping

Phosphorus-vacancy nearest-neighbor hopping (PVNNH) has been linked[13,14] to drain current drift (DCD) in InP metal-insulator-semiconductor field-effect transistors (MISFET's). Consider an InP MISFET or

MIS capacitor which has P vacancies in the channel region under its gate. Nearest-neighbor hopping of an In atom into the P vacancy can be described by the following defect reaction:

$$In_{In} + V_P^+ + 4e^- = V_{In}^- In_P^{-2} . \quad (11)$$

Application of the law of mass action to this defect reaction results in

$$\frac{[V_{In}^- In_P^{-2}]}{[V_P^+][In_{In}]} \alpha [e^-]^4 . \quad (12)$$

The application of a positive gate bias to an InP MIS capacitor or MISFET induces an accumulation of electrons in the channel and hence an increase in $[e^-]$ which shifts the equilibrium to the right-hand side of the defect reaction. Four of the accumulated channel electrons are captured for every P vacancy annihilated. DCD in MISFET's or flatband voltage shift in MIS capacitors is attributed to the loss of these electrons from the channel.

From the BM an activation enthalpy of $\Delta H_m(V_P) = 1.2$ eV was predicted[13] which was found to be in good agreement with experimental values[14] deduced from variable-temperature bias-stress measurements of InP MIS capacitors. Additionally, a computer simulation of the effect of the PVNNH mechanism on flatband voltage shift versus bias stress measurements was performed.[14] In the kinetic analysis it was found that the rate-limiting step in the total reaction given by Eq. (11) was

$$V_P^0 + e^- = (V_{In} In_P)^- \quad (13)$$

and an entropy of activation ΔS_a was used as an adjustable parameter (the only adjustable parameter) in the computer simulation to fit the experimental data. A value $\Delta S_a^{expt}/k_B = 15.3$ was deduced from the computer simulation[14] compared to that originally estimated[13] by Van Vechten and Wager from

$$\Delta S_a / k_B = \ln(8 D_0 / a^2 \nu_D) , \quad (14)$$

where D_0 is the prefactor of the diffusion constant for in self-diffusion in InP. Using an experimental value of $D_0 = 1 \times 10^5$ cm/s as reported[3] by Goldstein gives $\Delta S_a / k_B = 17.5$.

To calculate the entropy of activation we recognize that the rate-limiting reaction given by Eq. (13) involves both In hopping and ionization so that the activation entropy is given by

$$\Delta S_a = \Delta S_m + \Delta S_i \quad (15)$$

where ΔS_i is the entropy of ionization. We calculate ΔS_i using[16,18,23]

$$\Delta S_i(T) = \Delta S_{CV}(T) = \frac{\alpha T}{(T+\beta)^2} \quad (16)$$

where ΔS_{CV} is the entropy of the band gap (i.e., the entropy of formation of free-electron-hole pairs) and for InP (Ref. 24) $\alpha = 6.63 \times 10^{-4}$ eV/K and $\beta = 162$ K. ΔS_m is calculated using Eqs. (7)–(9) which results in an activation entropy $\Delta S_a^{theor}/k_B = 8.2 + 6.8 = 15.0$ (the parameters used in this calculation are summarized in Table I). In the calculation 325 K was used because it is the average temperature at which the activation energy was extracted in the computer simulation. We regard the close agreement between experiment (15.3) and theory (15.0) as strong support for the importance of the translational and ionization entropies in determining the total activation entropy.

B. The metastable *M* center

The metastable M center[8–12] is an electron-irradiation-induced defect complex in InP which exhibits two distinct configurations denoted A and B. There is a reversible transformation between the configurations which exhibits the following kinetics[8,9] as deduced from Arrhenius plots:

$A \rightarrow B$ stage 1:

$$k = 10^{18} \exp[-(0.40 \text{ eV})/k_B T], \; T = 110 \text{ K} \quad (17)$$

$A \rightarrow B$ stage 2:

$$k = 10^{11} \exp[-(0.42 \text{ eV})/k_B T], \; T = 160 \text{ K} \quad (18)$$

$B \rightarrow A$:

$$k = 10^{7} \exp[-(0.24 \text{ eV})/k_B T], \; T = 140 \text{ K} . \quad (19)$$

We have proposed[12] an atomic model for the M center. In terms of this atomic model, the reversible transformation $A = B$ may be written as follows:

$$(V_{In}^- P_{In}^+)^0 + 3e^- \underset{P \text{ hop}}{=} P_{In}^+ + V_P^+ + P_{In}^+ , \quad (20)$$

where the parentheses around the first term denotes a Coulombic attractive interaction between the initial point defects and P hop identifies P atomic migration to a nearest-neighbor vacancy as the transformation mechanism. We have discussed[12] the atomic model in some detail; it is our present objective to offer a quantitative explanation for the three widely divergent prefactors found in the kinetic equations.

First note that the reversible transformation between configurations as specified by Eq. (20) involves the capture or emission of three electrons as well as a P hop. We propose that the wide range of empirically deduced prefactors results from differences in the rate-limiting step in the defect complex transformation. We now individually treat each of the kinetic equations.

1. $A \rightarrow B$ stage 1

The process occurs at a temperature of 110 K with a prefactor $A_0^{expt}(110 \text{ K}) = 10^{18}$. We assume that the rate-limiting step is a process involving a P hop and emission of a single electron,

$$(V_{In}^- P_{In}^+)^0 \underset{P \text{ hop}}{\rightarrow} P_{In}^+ + V_P^0 + P_{In}^0 + e^- . \quad (21)$$

For this rate-limiting defect reaction the prefactor can be written as

$$A_0 = \nu_D \exp[(\Delta S_m + \Delta S_i)/k_B] . \quad (22)$$

The parameters used to calculate A_0 are summarized in Table I. The calculated prefactor $A(110\ \text{K}) = 1.2 \times 10^{18}$ is an excellent agreement with that found experimentally. Note the very large value of the activation entropy, $\Delta S_a / k_B = 12.2$ and that the translational entropy $S_t / k_B = 8.0$ is a dominant contribution.

2. $A \rightarrow B$ stage 2

This process occurs at 160 K with a prefactor $A_0^{\text{expt}}(160\ \text{K}) = 10^{11}\ \text{s}^{-1}$. We assume that the rate-limiting step is due to the emission of a single electron,

$$(V_{\text{In}}^{-} \text{P}_{\text{In}}^{+})^0 \rightarrow (V_{\text{In}}^{-} \text{P}_{\text{In}}^{+})^{+} + e^{-} . \tag{23}$$

The prefactor of the rate-limiting step corresponds to electron emission and can be written,

$$A_0 = \upsilon_{\text{th}} \sigma N_c \exp(\Delta S_i / k_B) , \tag{24}$$

where υ_{th} is thermal velocity, σ is the capture cross section, and N_c is the conduction-band effective density of states. Assuming $\sigma = 10^{-15}\ \text{cm}^2$, which is a typical capture cross section, we obtain $A_0^{\text{theor}}(160,\ \text{K}) = 9.6 \times 10^{11}\ \text{s}^{-1}$ (see Table I) again in good agreement with experiment.

3. $B \rightarrow A$

At a temperature of 140 K the prefactor of the $B \rightarrow A$ transformation is found to be $A_0^{\text{expt}}(140\ \text{K}) = 10^{7}\ \text{s}^{-1}$. In accordance[9] with Levinson *et al.* we assume that the reverse reaction occurs in one stage with the capture of three electrons as well as a P hop,

$$\text{P}_{\text{In}}^{+} + V_{\text{P}}^{+} + \text{P}_{\text{In}}^{+} + 3e^{-} \xrightarrow[\text{P hop}]{} (V_{\text{In}}^{-} \text{P}_{\text{In}}^{+})^0 . \tag{25}$$

The prefactor of this process can be written as,

$$A_0 = \nu_D \exp[(\Delta S_m - 3\Delta S_i)/k_B] \tag{26}$$

where the negative sign before the ionization entropy arises because of electron capture and a concomitant hardening of the lattice. As summarized in Table I, this re-

TABLE I. InP parameters used in the phosphorus vacancy nearest-neighbor hopping and metastable M center calculations.

	Lattice constant a (Å)	5.868 75
	Debye temperature θ_D (K)	292
	Debye frequency ν_D (s^{-1})	6.1×10^{12}
	Mass m (kg)	In $= 1.91 \times 10^{-25}$, P $= 5.14 \times 10^{-26}$
	Tetrahedral radius r (Å)	In $= 1.405$, P $= 1.128$
	Volume V (m^{-3})	1.76×10^{-29}
	Phosphorus vacancy nearest-neighbor hopping	
	Temperature T (K)	325
	Vibrational entropy S_v / k_B	3.4
	Translational entropy S_t / k_B	11.6
	Migrational entropy $\Delta S_m / k_B$	8.2
	Ionization entropy $\Delta S_i / k_B$	6.8
	Activation entropy $\Delta S_a / k_B$	15.0
	Metastable M center	
$A \rightarrow B$ stage 1		
	Temperature T (K)	110
	Vibrational entropy S_v / k_B	0.8
	Translational entropy S_t / k_B	8.0
	Migrational entropy $\Delta S_m / k_B$	7.2
	Ionization entropy $\Delta S_i / k_B$	5.0
	Activation energy $\Delta S_a / k_B$	12.2
	Prefactor A_0 (s^{-1})	1.2×10^{18}
$A \rightarrow B$ stage 2		
	Temperature T (K)	160
	Thermal velocity υ_{th} (cm/s)	3.1×10^{7}
	Capture cross section σ (cm^2)	10^{-15}
	Density of states N_c (cm^{-3})	1.04×10^{17}
	Ionization entropy $\Delta S_i / k_B$	5.7
	Prefactor A_0 (s^{-1})	9.6×10^{11}
$B \rightarrow A$		
	Temperature T (K)	140
	Vibrational entropy S_v / k_B	1.3
	Translational entropy S_t / k_B	8.4
	Migrational entropy $\Delta S_m / k_B$	7.1
	Ionization entropy $\Delta S_i / k_B$	−5.5
	Activation entropy $\Delta S_a / k_B$	−9.4
	Prefactor A_0 (s^{-1})	5.0×10^{8}

sults in a prefactor $A_0^{\text{theor}}(140\ \text{K})=4.0\times10^8\ \text{s}^{-1}$ in reasonable agreement with that found experimentally. Note that in this case, the large entropy of three captured electrons dominates, yielding a negative activation entropy.

C. Self-diffusion in Si and Ge

Let us now test our proposal for the entropy of atomic hopping by applying it to the experimentally deduced prefactors for self-diffusion in Ge and Si. The prefactor for self-diffusion D_0 for Ge over a temperature range of 766–928 °C was found to be[2]

$$D_0^{\text{expt}}(\text{Ge})=7.8\ \text{cm}^2/\text{s}\ . \tag{27}$$

The prefactor for Si self-diffusion when the self-diffusion data has been fit to a simple Arrhenius plot, has been reported[6] over a very wide range (1–9000 cm^2 s). Demond *et al.* have shown, however, that the rate of self-diffusion in Si does not obey[6] a simple Arrhenius law. They have interpreted this as due to the superposition of two activated processes. For the temperature range 830–1200 °C the dominant process has a prefactor that falls within the range,

$$D_0^{\text{expt}}(\text{Si,low}\ T)=0.2\text{–}20\ \text{cm}^2/\text{s}\ . \tag{28}$$

For temperatures greater than 1200 °C the dominant process has a prefactor of order

$$D_0^{\text{expt}}(\text{Si,high}\ T)\simeq2000\ \text{cm}^2/\text{s}\ . \tag{29}$$

We assume that self-diffusion in Ge and low-temperature self-diffusion in Si proceeds via neutral monovacancy diffusion. The prefactor for self-diffusion is given by[4,5,25]

$$D_0=\tfrac{1}{8}a^2 f\nu_D\exp[(\Delta S_m+\Delta S_f)/k_B]\ , \tag{30}$$

where a is the lattice constant, f is the correlation factor (equal to $\frac{1}{2}$ for self-diffusion by vacancy migration in diamond lattices[25]), and ΔS_f is the entropy of formation of a vacancy. ΔS_m is calculated using Eqs. (7)–(9). Note that ΔS_m is temperature dependent so that the average temperature over the experimental range is used to calculate ΔS_m. The parameters used to calculate ΔS_m for Si and Ge are summarized in Table II. The entropies of formation were taken[26,27] to be $\Delta S_f(\text{Si})/k_B=3.0$, $\Delta S_f(\text{Ge})/k_B=2.6$. Using Eq. (30), we find

$$D_0^{\text{theor}}(\text{Ge})=11\ \text{cm}^2/\text{s}\ , \tag{31}$$

$$D_0^{\text{theor}}(\text{Si,low}\ T)=24\ \text{cm}^2/\text{s} \tag{32}$$

which are in reasonable agreement with that found experimentally, particularly when it is realized that precise values for ΔS_f are not well established and there is a large spread in the experimental data in the case of Si self-diffusion. Note that the total activation entropy of the prefactor (i.e., $\Delta S_a=\Delta S_m+\Delta S_f$) is quite large for both Ge and Si, $\Delta S_a^{\text{theor}}(\text{Ge})/k_B=8.9$ and $\Delta S_a^{\text{theor}}(\text{Si})=9.2$.

An alternative mode of vacancy migration that yields the same theoretical prefactor for low-temperature Si self-diffusion is that singly ionized vacancies, rather than neutral vacancies, are the mediators of self-diffusion and that they are deionized through capture of free carriers concomitant with the hopping event, i.e.,

$$V_{\text{Si}}^{+}+e^{-}\xrightarrow[\text{hop}]{}V_{\text{Si}}^{x} \tag{33}$$

or

$$V_{\text{Si}}^{-}+h^{+}\xrightarrow[\text{hop}]{}V_{\text{Si}}^{x}\ . \tag{34}$$

The activation entropy for both of these self-diffusion mechanisms is the same as in Eq. (30), namely $(\Delta S_m+\Delta S_f)/k_B$, since the ionization entropy contribution from the singly ionized vacancy is canceled by the ionization entropy associated with the free-carrier capture. Note that Eqs. (33) and (34) for Si are similar to Eqs. (13), (21), and (25) for InP in that our prefactor analysis suggests that carrier capture or emission often occurs concomitant with atomic hopping.

The prefactor for high-temperature self-diffusion in Si

TABLE II. Parameters used to calculate the entropy of migration for Si and Ge self-diffusion. The values in parentheses refer to high-temperature self-diffusion in Si.

Parameter	Silicon	Germanium
Lattice constant a (Å)	5.43072	5.65754
Debye temperature θ_D (K)	648	374
Debye frequency ν_D (s^{-1})	1.35×10^{13}	7.79×10^{12}
Mass m (kg)	4.66×10^{-26}	1.21×10^{-25}
Tetrahedral radius r (Å)	1.173	1.225
Volume V (m^{-3})	1.35×10^{-29}	1.54×10^{-29}
Temperature T (K)	1288 (1573)	1120
Vibrational entropy S_v/k_B	5.1 (5.7)	6.3
Translational entropy S_t/k_B	11.3 (11.6)	12.6
Migrational entropy $\Delta S_m/k_B$	6.2 (5.9)	6.3
Formation entropy $\Delta S_f/k_B$	3.0	2.6
Ionization entropy $\Delta S_i/k_B$	(5.0)	
Activation entropy $\Delta S_a/k_B$	9.2 (13.9)	8.9
Prefactor D_0 (cm^2/s)	24 (2708)	11

can be justified in accordance with our activation entropy formulation if the total activation entropy used in Eq. (30) is given by

$$\Delta S_a = \Delta S_m + \Delta S_f + \Delta S_i \ . \tag{35}$$

Using an average temperature of 1573 K and Eq. (16) to evaluate ΔS_i with[16] $\alpha = 4.73 \times 10^{-4}$ eV/K and $\beta = 636$ K yields $\Delta S_i / k_B = 5.0$. This leads to (see Table II) $\Delta S_a / k_B = 13.9$ and a prefactor for high-temperature Si self-diffusion

$$D_0^{\text{theor}}(\text{Si, high } T) = 2708 \text{ cm}^2/\text{s} \tag{36}$$

which is in reasonable agreement with that deduced by Demond *et al.* There are various mechanisms for self-diffusion which would result in this prefactor: neutral vacancies which ionize concomitant with the hop to become singly ionized, singly ionized vacancies which hop without changing their ionization state, or doubly ionized vacancies which undergo free-carrier capture during the hop to become singly ionized. The situation becomes further complicated if the negative U character[28] of V_{Si} is brought in to consideration.

IV. CONCLUSIONS

We propose the entropy of atomic migration may be calculated from a difference of the translational and vibrational entropy as calculated using Eqs. (7)–(9). It is found that in a kinetic process involving atomic hopping is often dominated by the effects of the migrational and ionization entropies. We have been able to quantitatively account for the prefactors, or equivalently, the activation entropies, of PVNNH in InP MIS devices, transformations in the metastable M center in InP, and self-diffusion in Si and Ge. Prefactors with magnitudes as large as those exhibited in the processes discussed were previously considered to be anomalous. We believe that analysis of the magnitudes of prefactors, along the lines of that discussed herein, may be a powerful approach for the microscopic identification of defects and defect-related processes.

ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research, U. S. Department of Defense (DOD), under Contract Nos. AFOSR-86-0309 and AFOSR-89-0309.

*Present address: Tektronix, Inc., Beaverton, OR 97077.

[1]M. Boudart, *Kinetics of Chemical Processes* (Prentice-Hall, Englewood Cliffs, New Jersey, 1968).
[2]H. Letaw, W. M. Portnoy, and L. Slifkin, Phys. Rev. **102**, 636 (1956).
[3]B. Goldstein, Phys. Rev. **121**, 1305 (1961).
[4]A. Seeger and K. P. Chik, Phys. Status Solidi **38**, 3148 (1967).
[5]F. K. Kröger, *The Chemistry of Imperfect Crystals*, 2nd ed. (North-Holland, Amsterdam, 1974), pp. 292 and 294.
[6]F. J. Demond, S. Kalbitzer, H. Mannsperger, and H. Damjantschitsch, Phys. Lett. **93A**, 503 (1983).
[7]J. A. Van Vechten, Phys. Rev. B **33**, 8785 (1986).
[8]J. L. Benton and M. Levinson, Mater. Res. Soc. Symp. Proc. **14**, 95 (1983).
[9]M. Levinson, J. L. Benton, and L. C. Kimerling, Phys. Rev. B **27**, 6216 (1983).
[10]M. Levinson, M. Stavola, J. L. Benton, and L. C. Kimerling, Phys. Rev. B **28**, 5848 (1983).
[11]M. Stavola, M. Levinson, J. L. Benton, and L. C. Kimerling, Phys. Rev. B **30**, 832 (1984).
[12]J. F. Wager and J. A. Van Vechten, Phys. Rev. B **32**, 5251 (1985).
[13]J. A. Van Vechten and J. F. Wager, J. Appl. Phys. **57**, 1956 (1985).
[14]M. T. Juang, J. F. Wager, and J. A. Van Vechten, J. Electrochem. Soc. **135**, 2019 (1988); **135**, 2023 (1988).
[15]C. P. Flynn, *Point Defects and Diffusion* (Clarendon, Oxford, 1972).
[16]J. A. Van Vechten, in *Handbook of Semiconductors, Vol. 3*, edited by S. P. Keller (North-Holland, Amsterdam, 1980), Chap. 1.
[17]R. A. Swalin, J. Phys. Chem. Solids **18**, 290 (1961).
[18]J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B **14**, 3539 (1976).
[19]J. A. Van Vechten, Phys. Rev. B **12**, 1247 (1975).
[20]J. A. Van Vechten and J. F. Wager, Phys. Rev. B **32**, 5259 (1985).
[21]J. F. Lee, F. W. Sears, and D. L. Turcott, *Statistical Thermodynamics*, 2nd ed. (Addison-Wesley, Reading, Massachusetts, 1973).
[22]J. A. Van Vechten and J. C. Phillips, Phys. Rev. B **2**, 2160 (1970).
[23]P. A. Varotsos and K. D. Alexopoulos, *Thermodynamics of Point Defects and Their Relation with Bulk Properties* (North-Holland, Amsterdam, 1986).
[24]C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).
[25]C. Zener, J. Appl. Phys. **22**, 372 (1951).
[26]M. Lannoo and G. Allan, Phys. Rev. B **25**, 4089 (1982).
[27]J. A. Van Vechten, Mater. Res. Soc. Symp. Proc. **46**, 83 (1985).
[28]J. A. Van Vechten, Phys. Rev. B **33**, 2674 (1986).

Enthalpy of formation of antisite defects and antistructure pairs in III-V compound semiconductors

T. W. Dobson
Tektronix, Inc., Beaverton, Oregon 97077

J. F. Wager
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

(Received 30 March 1989; accepted for publication 11 May 1989)

A formulation for estimating the enthalpy of formation of neutral, isolated antisite defects in III-V compound semiconductors is presented, in which it is assumed that the formation enthalpy is comprised of disorder and electronic contributions. The disorder contribution was previously calculated by Van Vechten [J. Electrochem. Soc. **122**, 423 (1975); *Handbook on Semiconductors*, edited by S. P. Keller (North-Holland, Amsterdam, 1989), Vol. 3, Chap. 1]. The electronic contribution is equal to the difference of the donor (acceptor) levels with the valence-band maximum (conduction-band minimum). The enthalpy of formation of neutral, singly, and doubly ionized, isolated antisites is estimated for GaAs and InP. The enthalpy of formation of neutral antistructure pairs in wide band-gap III-V semiconductors is calculated assuming that the binding energy is given by the electrostatic energy of a doubly ionized nearest-neighbor pair of antisites separated by a normal lattice site. The enthalpy of formation of neutral antistructure pairs in the narrow band-gap III-V semiconductors InAs and InSb is estimated assuming the binding energy is twice the band-gap energy.

INTRODUCTION

The enthalpy of formation of a neutral antistructure pair $\Delta H_f(B_A A_B^x)$ is defined by the reaction

$$A_A B_B^x = B_A A_B^x; \quad \Delta H_f(B_A A_B^x), \tag{1}$$

corresponding to the interchange of two atoms in a unit cell of the crystal. Van Vechten has calculated[1,2] the enthalpies of formation of neutral antistructure pairs using the dielectric two-band model while treating the defect complex as a pair of impurities B_A and A_B. Within the context of this model, the enthalpy of formation of an antistructure pair is the sum of two terms,

$$\Delta H_f(B_A A_B) = \Delta H_0(B_A A_B) + \Delta H_e(B_A A_B). \tag{2}$$

The first term ΔH_0 is the contribution actually calculated[1,2] by Van Vechten using the dielectric two-band model, and is a disorder term associated with the reduction in the optical band gap resulting from the disorder concomitant with the formation of antisite defects. The second term is an electronic contribution due to the ionization charge state of the defect complex. Van Vechten has argued[1,2] that since an antistructure pair is neutral within the unit cell, the electronic contribution will be zero, and the enthalpy of formation of an antistructure pair is entirely due to the first term,

$$\Delta H_f(B_A A_B^x) = \Delta H_0(B_A A_B^x). \tag{3}$$

Values for $\Delta H_0(B_A A_B^x)$ and thus for $\Delta H_f(B_A A_B^x)$ as calculated[1,2] by Van Vechten are given in Table I.

The first goal of the work discussed in this paper is to estimate the enthalpy of formation of isolated antisite defects in III-V compound semiconductors. To accomplish this, we first estimate the enthalpy of formation of isolated, neutral antisite defects, and from these values infer the enthalpy of formation of the neutral antistructure pair. Note that this approach differs[1,2] from that originally taken by Van Vechten, who began by considering the antistructure pair formation enthalpy and then inferred the formation enthalpy of the individual antisites. Thus, the second goal of this paper is to re-estimate the enthalpy of formation of antistructure pairs in III-V compound semiconductors.

ENTHALPY OF FORMATION OF ANTISITE DEFECTS

Isolated antisites are not necessarily electrically neutral. An A_B antisite defect in a III-V compound semiconductor exists as a double acceptor[3] with a possible charge state of 0, -1, or -2, while a B_A antisite is a double donor with a charge state of 0, $+1$, or $+2$. The actual charge state in which an antisite exists depends on the location in energy of

TABLE I. Disorder contribution to the enthalpy of formation of antisite defects and antistructure pairs in III-V compound semiconductors. Enthalpy is expressed in units of eV.

Semiconductor	$\Delta H_0(B_A A_B^x)$	$\Delta H_0(A_B^x)$	$\Delta H_0(B_A^x)$
AlP	1.10	0.60	0.50
AlAs	0.75	0.30	0.45
AlSb	0.38	0.02	0.36
GaP	1.06	0.68	0.38
GaAs	0.70	0.35	0.35
GaSb	0.40	0.08	0.32
InP	1.11	0.89	0.42
InAs	0.90	0.57	0.33
InSb	0.54	0.27	0.27

 0021-8979/89/171997-05$02.40

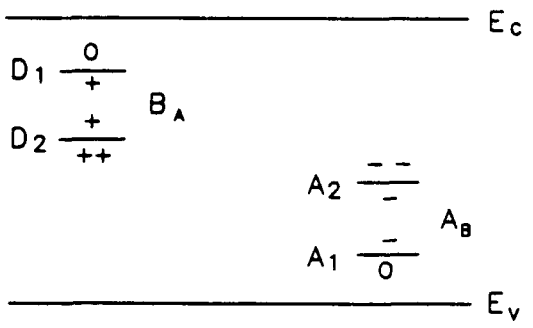


FIG. 1. An energy-band diagram which illustrates that the electronic contribution to the enthalpy of formation of a neutral anion antisite defect is equal to the difference in the donor levels with respect to the valence-band maximum, and is equal to the difference in the acceptor levels with respect to the conduction-band minimum for a neutral cation antisite.

the ionization levels and on the position of the Fermi level. In general, the enthalpy of formation of an antisite defect will have both disorder and electronic contributions,

$$\Delta H_f(A_B^{-z}) = \Delta H_0(A_B^x) + \Delta H_e(A_B^{-z}), \quad (4)$$

$$\Delta H_f(B_A^{+z}) = \Delta H_0(B_A^x) + \Delta H_e(B_A^{+z}), \quad (5)$$

where $z = 0$, 1, or 2 is the charge state of the antisite defect.

Van Vechten has previously calculated[1,2] the disorder contribution to the formation enthalpy for isolated, neutral antisite defects. These values are presented in Table I for III-V compound semiconductors. To determine the total enthalpy of formation, it is necessary to calculate the electronic contribution ΔH_e.

The electronic contribution ΔH_e for an isolated anion antisite is calculated as follows. Consider a neutral anion atom sitting on its normal lattice site B_B^x. For the purpose of calculating $\Delta H_e(B_A^x)$, we envisage B_B^x as a neutral donor with ionization levels that lie at the valence-band maximum E_v. The neutral anion antisite B_A^x, which is the analog of B_B^x, possesses two ionization levels E_{D1} and E_{D2}, which are within the energy band gap, as shown in Fig. 1. Thus, replacement of B_B^x by B_A^x requires an electronic energy equal to the difference of the donor levels with respect to the valence-band maximum. The electronic contribution to the enthalpy of formation of a neutral anion antisite is simply

$$\Delta H_e(B_A^x) = (E_{D1} - E_V) + (E_{D2} - E_V), \quad (6)$$

where E_{D1} and E_{D2} are the first- and second-donor ionization energies, respectively. Similarly, the electronic contribution to the enthalpy of formation of a neutral cation antisite is given by

$$\Delta H_e(A_A^x) = (E_C - E_{A1}) + (E_C - E_{A2}), \quad (7)$$

where E_C denotes the bottom of the conduction band, and E_{A1} and E_{A2} are the first and second ionization energies, respectively.

It is now possible to calculate the enthalpy of formation of neutral, isolated antisite defects from Eqs. (4) and (5) with the help of Eqs. (6) and (7) for evaluation of the electronic contribution and, with the aid of Table I, for evaluation of the disorder contribution. It is possible, however, that the antisite defect is singly or doubly ionized instead of neutral; the Fermi-level position and ionization energies of the antisite will determine the actual charge state. If the antisite is ionized and if the ionization energies of the antisite are known for the semiconductor of interest, the enthalpy of formation of the antisite in other charge states may be calculated using the Law of Mass Action to give [see the Appendix for a derivation of Eq. (8)]

$$\Delta H_f(B_A^+) = \Delta H_f(B_A^{+2}) + (E_{D2} - E_F), \quad (8)$$

$$\Delta H_f(B_A^+) = \Delta H_f(B_A^x) + (E_F - E_{D1}), \quad (9)$$

$$\Delta H_f(A_B^-) = \Delta H_f(A_B^{-2}) + (E_F - E_{A2}), \quad (10)$$

$$\Delta H_f(A_B^-) = \Delta H_f(A_B^x) + (E_{A1} - E_F), \quad (11)$$

where E_F is the Fermi level.

The enthalpies of formation of isolated antisite defects of various charge states are given in Table II for GaAs and InP. These values are calculated using the ionization energies given in Table III and assuming a Fermi-level position at the band edges E_C or E_V, or at midgap E_I. Note that we have employed ionization energies which are estimated for $T = 0$ K in this work; we ignore the temperature dependence of these ionization energies.

It is apparent from Tables I and II that the electronic term ΔH_e is a major component of the total enthalpy of formation of individual antisite defects for III-V compound semiconductors. It is also important to note that the Fermi-

TABLE II. Enthalpies of formation of isolated antisite defects of various charge states for (a) GaAs and (b) InP. These values were calculated for three Fermi-level positions and by assuming that the antisite ionization energies are those given in Table III. Enthalpy is expressed in units of eV.

(a) GaAs						
E_F position	$\Delta H_f(B_A^x)$	$\Delta H_f(B_A^+)$	$\Delta H_f(B_A^{+2})$	$\Delta H_f(A_B^x)$	$\Delta H_f(A_B^-)$	$\Delta H_f(A_B^{-2})$
E_C	1.62	2.39	3.39	3.11	1.67	0.35
E_I	1.62	1.63	1.87	3.11	2.43	1.87
E_V	1.62	0.87	0.35	3.11	3.19	3.39
(b) InP						
E_C	2.38	2.39	3.26	3.57	2.28	0.89
E_I	2.38	1.68	1.84	3.57	2.99	2.31
E_V	2.38	0.97	0.42	3.57	3.70	3.73

TABLE III. Estimated ionization energies at $T = 0$ K of isolated antisite defects in GaAs (Ref. 4) and InP (Ref. 5). Energies are with respect to the conduction band minimum for donors and the valence band maximum for acceptors.

Defect	Ionization level	Charge state	Ionization Energy (eV) GaAs	InP
A_B	$A1$	$-/0$	0.078	0.03
	$A2$	$-/--$	0.20	0.13
B_A	$D1$	$0/+$	0.77	0.01
	$D2$	$+/++$	1.00	0.87

level position and the ionization energy positions have a profound effect on the energy of formation of antisite defects, particularly in the case of wide band-gap semiconductors such as GaAs and InP. Even when the electronic contribution is properly taken into account, thermodynamic energy accounting leads to the conclusion that antisites and antistructure pairs typically have a smaller energy of formation than vacancies and divacancies. Thus, antisites should be anticipated as common defects in III-V compond semiconductors.

ENTHALPY OF FORMATION OF ANTISTRUCTURE PAIRS

The neutral antistructure pair $B_A A_B^x$ may be envisaged as a nearest-neighbor pair comprised of A_B^{-z} and B_A^{+z}, where z is the charge state of the individual antisites. We argue that $z = 2$ for both components of the antistructure pair, independent of the ionization energy locations of the individual antisite defects as long as the Fermi level is within approximately the middle half of the band gap. This follows from the effect of associates of oppositely charged defects on the position of the ionization level.[3] If a positively charged defect combines with negatively charged defect to form a pair, the ionization energies of this defect complex differ from that of the individual isolated defects; the donor levels move toward the conduction band and the acceptor levels toward the valence band. The total change in the ionization level positions due to association can be approximated[3] as the electrostatic energy due to two nearest-neighbor point charges,

$$\Delta E = Z_A Z_B q^2/4\pi\epsilon r, \tag{12}$$

where Z_A and Z_B are the charges of the two antisite defects comprising the antistructure pair, r is the equilibrium nearest-neighbor separation distance, and ϵ is the dielectric constant. For III-V compound semiconductors, $Z_A = Z_B = 2$, $\epsilon \simeq 12$, and $r \simeq 2.5$ Å which gives $\Delta E \simeq 1.8$ eV. This is approximately equal to, and in some cases much greater than, the band gap of the III-V semiconductor under consideration. Thus, the effect of association is to force the ionization levels close to or resonant with the conduction or valence bands so that the individual antisites attain their full ionization state of 2, and the antistructure pair is net neutral $B_A^{+2} A_B^{-2} = B_A A_B^x$.

The binding energy of the neutral antistructure pair $\Delta H_b(B_A A_B^x)$ is defined as the energy required to separate it into two isolated antisite defects,

$$B_A A_B^x = B_A^{+2} + A_B^{-2}; \quad \Delta H_b(B_A A_B^x). \tag{13}$$

We assume that this binding energy is approximated by some fraction α of the electrostatic energy of the doubly ionized antisite defects occupying unperturbed lattice site positions,

$$\Delta H_b(B_A A_B^x) = 4q^2\alpha/4\pi\epsilon r. \tag{14}$$

Some discussion of Eq. (14) is required. Within the framework of our assertions, the binding energy of a neutral antistructure pair is identified as the Coulomb interaction energy of the doubly ionized, nearest-neighbor antisite defect pair as estimated in the fully screened, point charge approximation. Additionally, this interaction energy is identified as the energy of association which forces the ionization levels near to or resonant with the respective band edges. We assert that the fully screened, point charge approximation is a reasonable estimate of the antistructure pair binding energy as long as the antistructure pair ionization energies are present within the band gap and are, thus, localized electronic states. When the ionization levels become resonant with the band edge, however, the fully screened, point charge approximation is of questionable validity since these states are now resonant and, hence, nonlocal in nature.

Without becoming entrenched in these subtleties, we postulate that only a fraction α of the Coulombic energy contributes to the antistructure pair binding energy; this fraction corresponds to the amount of association energy it takes to push the antistructure ionization levels out of the band gap. In order to evaluate α, we must know the ionization energies of the isolated antisite defects as well as the fraction of energy which goes into modifying each ionization level. It is clear that α will be greater for wide band-gap semiconductors compared to narrow band-gap semiconductors like InAs and InSb, since a small fraction of the association energy is sufficient to push the ionization levels resonant with the band edges for semiconductors with small band gaps.

In order to estimate the enthalpy of formation of an antistructure pair in a III-V semiconductor without knowing α, we adopt the following approach. We assume the antistructure binding energy to be the smaller of: (a) that calculated from Eq. (14) with $\alpha = 1$ or (b) twice the band-gap energy $2E_{CV}$ which is the association energy required to push four ionization levels from midgap to the band edges. Thus,

$$\Delta H_b^{\max}(B_A A_B^x) = 4q^2/(4\pi\epsilon r) \text{ or } 2E_{CV}, \tag{15}$$

whichever is smaller. As noted in Eq. (15), this estimate corresponds to a maximum value of the binding energy which would, thus, correspond to a minimum value of the enthalpy of formation of an antistructure pair.

With these issues resolved, we can return to the problem of estimating the enthalpy of formation of a neutral antistructure pair, where we will now employ Eq. (15) as an estimate of the binding energy,

$$\Delta H_f(B_A A_B^x) = \Delta H_f(A_B^{-2}) + \Delta H_f(B_A^{+2}) - \Delta H_b(B_A A_B^x). \quad (16)$$

From Eqs. (4), (7), (10), and (11),

$$\Delta H_f(A_B^{-2}) = \Delta H_0(A_B^x) + 2(E_C - E_F) \quad (17)$$

and from Eqs. (5), (6), (8), and (9),

$$\Delta H_f(B_A^{+2}) = \Delta H_0(B_A^x) + 2(E_F - E_V). \quad (18)$$

Combining Eqs. (16), (17), and (18) leads to

$$\Delta H_f(B_A A_B^x) = \Delta H_0(B_A A_B^x) + 2E_{CV} - \Delta H_b(B_A A_B^x), \quad (19)$$

where E_{CV} is the band gap. The enthalpy of formation of a neutral antistructure pair may be estimated using Eqs. (15) and (19) without any knowledge of the ionization energies of the defect complex. These values are compiled in Table IV for III-V semiconductors. Our values for 1.95 and 1.79 eV for the antistructure pair enthalpy of formation and binding energy, respectively, are in rough agreement with that calculated[6] by Baraff and Schluter ($\sim$1.7 and $\sim$2.5 eV).[7]

It can be seen from Table IV that the enthalpy of formation of an antistructure pair is greater[1,2] than that originally estimated by Van Vechten for all of the III-V semiconductors considered, except for InAs and InSb, which are identical to his original estimates. In his original formulation, Van Vechten considered only the disorder contribution to the enthalpy of formation of an antistructure pair, whereas in this work we also include electronic and binding contributions. Our formulation of the antistructure pair enthalpy of formation as given by Eq. (19) reduces to that of Van Vechten when it is assumed that the electronic and binding contributions are of equal magnitude. Note from Table IV that this assumption is employed only for the small band-gap semiconductors InAs and InSb, in which case the Coulombic energy is greater than twice the band gap. For all of the other III-V semiconductors considered, the Coulombic energy is less than twice the band gap, so that the binding energy is estimated by the fully screened, point charge approximation. Note once again that the antistructure pair enthalpies of formation listed in Table IV represent lower limit estimates since Eq. (15) is used to estimate the binding energy.

TABLE IV. Enthalpies of formation of antistructure pairs in III-V compound semiconductors. $\Delta H_0(B_A A_B)$ is the disorder contribution, E_{CV} is the low-temperature band gap, and ΔH_b is the enthalpy of binding of the antistructure pair as estimated from Eq. (15). All quantities are expressed in units of eV.

Semiconductor	$\Delta H_0(B_A A_B^x)$	E_{CV}	ΔH_b	$\Delta H_f(B_A^{+2} A_B^{-2})$
AlP	1.10	2.52	2.49	3.65
AlAs	0.75	2.25	2.15	3.10
AlSb	0.38	1.68	1.50	2.24
GaP	1.06	2.34	2.20	3.54
GaAs	0.70	1.52	1.79	1.95
GaSb	0.40	0.81	1.39	0.63
InP	1.11	1.42	1.83	2.12
InAs	0.90	0.42	0.84	0.90
InSb	0.54	0.23	0.46	0.54

ACKNOWLEDGMENTS

We wish to thank J. A. Van Vechten for useful discussions. This work was supported by the Air Force Office of Scientific Research under Contract Nos. AFOSR 86-0309 and 89-0309.

APPENDIX

A derivation of the relation between the enthalpy of formation of a fully and partially ionized antisite is given for the case of a doubly ionized anion antisite which becomes singly ionized,

$$B_A^{+2} + e^- = B_A^+. \quad (A1)$$

Application of the Law of Mass Action leads to

$$K = [B_A^+]/[B_A^{+2}][e^-] \quad (A2)$$

where K is the equilibrium constant and the square brackets denote the concentration of the enclosed species. From semiconductor statistics considerations,[7]

$$[e^-] = N_c \exp[(E_F - E_C)/kT], \quad (A3)$$

where N_c is the conduction-band effective density of states, E_C is the energy of the bottom of the conduction band, and k is Boltzmann's constant. Also, from basic semiconductor statistics,[7]

$$[B_A^+]/[B_A^{+2}] = g_{D2} \exp[(E_F - E_{D2})/kT], \quad (A4)$$

where g_{D2} and E_{D2} are the degeneracy and energy, respectively, of the second ionization level. Substitution of Eqs. (A3) and (A4) into (A2) leads to

$$K = g_{D2}/N_c \exp[(E_C - E_{D2})/kT]. \quad (A5)$$

The Gibbs free energy ΔG_R of the defect reaction specified in Eq. (A1) is given by

$$\Delta G_R = \Delta G_f(B_A^+) - \Delta G_f(B_A^{+2}) - \Delta G_f(e^-), \quad (A6)$$

where ΔG_f denotes the Gibbs free energy of formation. The Gibbs free energy of reaction is related to the equilibrium constant via

$$\Delta G_R = -kT \ln K. \quad (A7)$$

Substitution of Eqs. (A5) and (A6) into (A7) gives

$$\Delta G_f(B_A^+) - \Delta G_f(B_A^{+2}) - \Delta G_f(e^-) = -kT \ln g_{D2}/N_c - (E_C - E_{D2}). \quad (A8)$$

The Gibbs free energy of the free electron is given by

$$\Delta G_f(e^-) = E_C - E_F, \quad (A9)$$

which when substituted in to Eq. (A8) leads to

$$\Delta G_f(B_A^+) = \Delta G_f(B_A^{+2}) + (E_{D2} - E_F) - kT \ln g_{D2}/N_c. \quad (A10)$$

The last term in Eq. (A10) is an entropy term.[8] Also recall that

$$\Delta G_f = \Delta H_f - T\Delta S_f, \quad (A11)$$

which leads to the desired result when the enthalpy and entropy terms substituted into Eq. (A10) are equated.

$$\Delta H_f(B_A^+) = \Delta H_f(B_A^{+2}) + (E_{D2} - E_F). \quad \text{(A12)}$$

Equation (A12) is equivalent to Eq. (8). Equations (9)–(11) may be derived in a similar manner.

[1] J. A. Van Vechten, J. Electrochem. Soc. **122**, 423 (1975).
[2] J. A. Van Vechten, in *Handbook on Semiconductors*, edited by S. P. Keller (North-Holland, Amsterdam, 1989), Vol. 3, Chap. 1.
[3] F. A. Kroger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1974), Vol. 2, p. 134.
[4] J. F. Wager and J. A. Van Vechten, Phys. Rev. B **35**, 2330 (1987).
[5] J. A. Van Vechten and J. F. Wager, J. Appl. Phys. **57**, 1956 (1985).
[6] G. A. Baraff and M. Schluter, Phys. Rev. B **33**, 7346 (1986).
[7] J. S. Blakemore, *Semiconductor Statistics* (Dover, New York, 1962).
[8] O. Engstrom and A. Alm, Solid-State Electron. **21**, 1571 (1978).

IV. CONCLUSIONS

An atomistic thermodynamic formulation for the energetics of self-diffusion in GaAs is presented. Self-diffusion involving vacancy migration modes is given primary emphasis, although Frenkel pair formation is also briefly considered. Our list of vacancy migration mechanisms is not exhaustive; for example, we do not treat self-diffusion by trivacancies as proposed[45] by Van Vechten and Schmid. We conclude that if vacancies are available in equilibrium concentrations (in other words, if local equilibrium is maintained by the creation of vacancies at nearby surfaces and dislocations), vacancy 2nnh is always predicted to be the most favorable mechanism of self-diffusion. Self-diffusion by vacancy 2nnh is compatible with prefactors of magnitude $\sim 10^{-5}$–10^{-1} $cm^2 s^{-1}$ and activation energies $\sim$3–4 eV. The energetics of self-diffusion depends strongly on the Fermi-level position and thus on the charge states and ionization energies of the defect complexes involved in self-diffusion. When the Fermi level is near the conduction-band minimum, Ga diffusion by 2nnh is predicted to dominate. In contrast, As diffusion by 2nnh is predicted to have the minimum ΔH_A when the Fermi level is near the valence-band maximum. When the Fermi level is near the middle of the bandgap, 2nnh by Ga and As are almost equivalent energetically and possess the smallest ΔH_A's of the mechanisms considered. These observations are consistent with certain self-diffusion trends as reviewed[9] by Deppe and Holonyak.

It is proposed that Ga self-diffusion or interdiffusion processes which are characterized by large prefactors ($\sim 10^7$–10^8 $cm^2 s^{-1}$) and activation energies ($\sim$6 eV) involve cation Frenkel pair formation. This mode of self-diffusion is expected to be dominant when surface vacancy generation is precluded and vacancies can only be generated in the bulk of the semiconductor by the creation of an interstitial-vacancy pair. Although nnh can also account for very large prefactors and activation energies, we prefer to invoke Frenkel pair formation because our energetic estimation indicates 2nnh to be always more energetically favorable than nnh if vacancies are available to mediate self-diffusion.

ACKNOWLEDGMENTS

I wish to thank J. A. Van Vechten for several informative discussions during the course of this work. This work was supported by the Air Force Office of Scientific Research under Contract No. 89-0309.

[1] B. Goldstein, Phys. Rev. **121**, 1305 (1961).
[2] H. R. Potts and G. L. Pearson, J. Appl. Phys. **37**, 2098 (1966).
[3] S. Y. Chiang and G. L. Pearson, J. Appl. Phys. **46**, 2986 (1975).
[4] D. L. Kendall, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1968), Vol. 4 Chap. 3.
[5] J. S. Blakemore, J. Appl. Phys. **53**, 520 (1982).
[6] H. D. Palfrey, M. Brown, and A. F. W. Willoughby, J. Electrochem. Soc. **128**, 2224 (1981).
[7] H. D. Palfrey, M. Brown, and A. F. W. Willoughby, J. Electron. Mater. **12**, 863 (1983).
[8] A. F. W. Willoughby, in *Defects in Semiconductors II*, Mater. Res. Soc. Symp. Proc., Vol. 14, edited by S. Mahajan and J. W. Corbett (Elsevier, New York, 1983), p. 237.
[9] D. G Deppe and N. Holonyak, Jr., J. Appl. Phys. **64**, R93 (1988).
[10] T. Y. Tan and U. Gösele, Appl. Phys. Lett. **52**, 1240 (1988).
[11] P. M. Petroff, J. Vac. Sci. Technol. **14**, 973 (1977).
[12] R. M. Fleming, D. B. McWhan, A. C. Gossard, W. Wiegmann, and R. A. Logan, J. Appl. Phys. **51**, 357 (1980).
[13] J. Cibert, P. M. Petroff, D. J. Werder, S. J. Pearson, A. C. Gossard, and J. H. English, Appl. Phys. Lett. **49**, 223 (1986).
[14] T. E. Schlesinger and T. Kuech, Appl. Phys. Lett. **49**, 519 (1986).
[15] P. Mei, H. W. Yoon, T. Venkatesan, S. A. Schwartz, and J. P. Harbison, Appl. Phys. Lett. **50**, 1823 (1987).
[16] P. Mei, S. A. Schwartz, T. Venkatesan, C. L. Schwartz, and E. Colas, Mater. Res. Soc. Symp. Proc. **126**, 71 (1988).
[17] T. Y. Tan and U. Gösele, Mater. Res. Soc. Symp. Proc. **144**, 21 (1988).
[18] L. J. Guido, N. Holonyak, K. C. Hsieh, R. W. Kaliski, W. E. Plano, R. D. Burnham, R. L. Thornton, J. E. Epler, and T. L. Paoli, J. Appl. Phys. **61**, 1372 (1987).
[19] J. D. Ralson, S. O'Brien, G. W. Wicks, and L. F. Eastman, Appl. Phys. Lett. **52**, 1511 (1988).
[20] J. A. Van Vechten, in *Handbook on Semiconductors*, edited by S. P. Keller (North-Holland, Amsterdam, 1980), Vol. 3, Chap. 1.
[21] J. F. Wager and J. A. Van Vechten, Phys. Rev. B **35**, 2330 (1987)
[22] J. A. Van Vechten, J. Electrochem. Soc. **122**, 419 (1975); **122**, 423 (1975).
[23] T. W. Dobson and J. F. Wager, J. Appl. Phys. **66**, 1997 (1989).
[24] J. A. Van Vechten, Phys. Rev. B **8**, 3351 (1970).
[25] J. A. Van Vechten, Phys. Rev. B **11**, 3910 (1975).
[26] J. A. Van Vechten, Phys. Rev. B **12**, 1247 (1975).
[27] G. A. Baraff and M. Schlüter, Phys. Rev. B **33**, 7346 (1986).
[28] F. A. Kroger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1964) (revised 1974).
[29] G. A. Baraff and M. Schlüter, Phys. Rev. Lett. **55**, 1327 (1986).
[30] D. C. Look, D. C. Walters, and J. R. Meyer, Solid State Commun. **42**, 745 (1982).
[31] E. R. Weber, H. Ennen, U. Kaufmann, J. Windscheif, J. Schneider, and T. Wosinski, J. Appl. Phys. **53**, 6140 (1982).
[32] M. Bugajski, K. H. Ko, J. Lagowski, and H. C. Gatos, J. Appl. Phys. **65**, 596 (1988).
[33] T. W. Dobson, J. F. Wager, and J. A. Van Vechten, Phys. Rev. B **40**, 2962 (1989).
[34] J. F. Wager, Philos. Mag. A (in press).
[35] J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B **14**, 3539 (1976).
[36] C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).
[37] M. Lannoo and G. Allan, Phys. Rev. B **25**, 4089 (1982).
[38] J. A. Van Vechten, Phys. Rev. B **33**, 8785 (1986).
[39] M. Lannoo and G. Allan, Phys. Rev. B **33**, 8789 (1986).
[40] L. J. Guido, N. Holonyak, Jr., K. C. Hsieh, and J. E. Baker, Appl. Phys. Lett. **54**, 262 (1989).
[41] A. Ourmazo, Y. Kim, and M. Bode, Mater. Res. Soc. Symp. Proc. **163**, 639 (1990).
[42] The author is indebted to the reviewer of this paper for suggesting that self-diffusion with $E_a \simeq 6$ eV is likely due to bulk generation via Frenkel pairs.
[43] D. G. Deppe, N. Holonyak, Jr., W. E. Plano, V. M. Robbins, J. M. Dallesasse, K. C. Hsieh, and J. E. Baker, J. Appl. Phys. **64**, 1838 (1988)
[44] S. Fujita, J. Phys. Chem. Solids **49**, 41 (1988).
[45] J. A. Van Vechten and U. Schmid, J. Vac. Sci. Technol. B **7**, 827 (1989).

Now consider the entropies of activation $\Delta S_a/k_B = 6.2, -3.4$ as determined[4,6-8] by Kendall and Palfrey *et al.*, respectively. As mentioned above, $\Delta S_a/k_B \simeq 6$ for mechanism 1, Ga diffusion by 2nnh which is in good agreement with that reported[4] by Kendall but is significantly larger than that deduced[6-8] by Palfrey *et al.* The essential point, however, is that the activation entropy for 2nnh is expected to be much less than that for nnh.

IV. DISCUSSION

A. Vacancy-mediated self-diffusion

The energetics of self-diffusion in GaAs has been discussed in terms of atomistic thermodynamic modeling of vacancy modes of self-diffusion. The activation enthalpy analysis led to the conclusion that, if vacancies are available in thermodynamic equilibrium concentrations, self-diffusion should occur by 2nnh. Our assessment of the energetics of 2nnh is consistent with reported activation energies of ~3–4 eV and is compatible with As-rich, *n*-type QWH annealing trends in which intermixing is attributed to cation vacancy 2nnh. Our analysis predicts As-poor, *p*-type QWH intermixing to be due to As 2nnh; this is clearly not the case because intermixing is found to occur on the cation sublattice.

Additionally, an activation entropy analysis has been performed. The primary conclusion of this analysis is that "normal" activation entropies (i.e., entropies associated with prefactors of order 10^{-5}–10^{-1} $cm^2\,s^{-1}$ and with activation energies ~3–4 eV) are compatible with 2nnh whereas "abnormally large" activation entropies (i.e., entropies associated with prefactors of order 10^7–10^8 $cm^2\,s^{-1}$ and with activation energies ~6 eV) are more compatible with nnh.

When the conclusions obtained from the activation enthalpy and entropy analyses are considered together with experimentally established self-diffusion and QWH annealing trends, it is tempting to identify large prefactors ($\sim 10^7$–10^8 $cm^2\,s^{-1}$) and large activation energies (~6 eV) as arising from nnh whereas small prefactors ($\sim 10^{-5}$–10^{-1} $cm^2\,s^{-1}$) and small activation energies (~3–4 eV) as due to 2nnh. There is a problem, however, with this identification. From Table V it is difficult to understand why diffusion should ever occur by nnh around a sixfold ring because of the unfavorable energetics. The primary point of Table V is that *if vacancies are available in equilibrium concentrations* the energetic assessment favors vacancy 2nnh as the dominant mode of self-diffusion.

B. Interstitial-mediated self-diffusion

The key issue implicitly assumed in the previous analysis is that vacancies are available in equilibrium concentrations. Another assumption implicit in the vacancy formation energetic assessment is that vacancies are created at nearby surfaces or dislocations and exist in a state of local equilibrium. Recent work[40,41] by Guido *et al.* and Ourmazo *et al.*, for example, indicate that this surface-vacancy creation, local equilibrium assumption is often invalid and that interdiffusion is strongly depth dependent due to the lack of native point defects in the bulk required to mediate self-diffusion.

With this in mind we propose[42] that in self-diffusion or interdiffusion situations in which diffusion occurs in a bulk region remote from surfaces and dislocations, a new vacancy can be created only as a Frenkel pair, i.e., a self-interstitial-vacancy pair. Self-diffusion by Frenkel pair formation (i.e., an enhanced solubility of the cation self-interstitial) has previously been invoked[9,43] by Deppe *et al.* to account for QWH interdiffusion trends in which intermixing occurs under *p*-type, As-rich conditions. Employing Van Vechten's method[20] for estimating the formation enthalpy of such a neutral Frenkel pair,

$$\Delta H_f(A_i^x \text{ or } B_i^x) = E_g(AB)Z(A_i \text{ or } B_i)/2, \tag{21}$$

where A_i and B_i refer to a cation or anion self-interstitial, respectively, E_g refers to the dielectrically averaged band gap, and Z is the valence of the element which becomes the self-interstitial. This leads to $\Delta H_f(Ga_i^x) = 7.8$ eV and $\Delta H_f(As_i^x) = 13.0$ eV. Note that Eq. (21) is not expected to hold better than ±20%.[20] Thus, according to this estimate it is feasible that self-diffusion or interdiffusion experiments with measured activation energies of ~6 eV are dominated by cation Frenkel pair formation. In other words, deep in the bulk Ga vacancies are created by Frenkel pair generation. The generated V_{Ga} would then be available to mediate self-diffusion, presumably by Ga 2nnh. The large energy associated with the formation of the self-interstitial would preclude a Ga_i from existing in an equilibriumlike manner so that it would be expected to be highly mobile and, thus, it could travel large distances before it is annihilated at a vacancy. Large diffusion distances would also be expected because of the small trapping cross section of an interstitial by a vacancy due to the large amount of energy that needs to be dissipated in the interstitial-vacancy annihilation.

Note that the energy of formation of an anion Frenkel pair is exceedingly large, according to the theory[20] of Van Vechten.

Before it can be confidently concluded that Frenkel pair formation is indeed the mechanism responsible for self-diffusion and interdiffusion processes with activation energies of ~6 eV, it is necessary to justify that Frenkel pair mediated self-diffusion can account for the associated abnormally large prefactor of ~ 10^8 $cm^2\,s^{-1}$. Although this issue remains to be resolved in detail we note that Fujita has developed[44] a statistical mechanical theory of fast interstitial diffusion in metals and he discusses prefactors of abnormally large magnitude. Specifically, he attributes $D_0 = 10^6$ $cm^2\,s^{-1}$ for Kr diffusion in KCl to interstitial diffusion of neutral Kr. In analogy to his identification, interstitial diffusion of neutral self-interstitials could give rise to prefactors of abnormally large magnitude.

$\Delta S_a/k_B = 6.2$ whereas $D_0 = 4.6 \times 10^{-5}$ cm^2 s^{-1} as measured[6-8] by Palfrey *et al.* yields $\Delta S_a/k_B = -3.4$.

Assessment of ΔS_a is accomplished using the formulation[33,34] of Dobson *et al.* and Wager. In general, the activation entropy can be envisaged as a sum of three contributions,

$$\Delta S_a = \Delta S_m + \Delta S_i + \Delta S_f, \tag{11}$$

where ΔS_m, ΔS_i, and ΔS_f are, respectively, the entropy of migration, ionization, and formation. The entropy of migration is given by[33,34]

$$\Delta S_m = S_s - S_v, \tag{12}$$

where S_s is the saddle-point entropy of the atom undergoing hopping and is given by

$$S_s = \frac{1}{4} k_B + k_B \ln\left[\frac{1}{4^{5/4}\sqrt{2\pi}}\left(\frac{k_B T}{\mathrm{KE}_m}\right)^{1/4} \times \left(\frac{(2\pi e m k_B T)^{3/2} V}{h^3}\right)\right], \tag{13}$$

where $e = 2.718$, m is the mass of the hopping atom, V is the hopping volume which, for nnh, we take to be the volume occupied by two nearest neighbors, and KE_m is the minimum kinetic energy required for hopping according to the ballistic model.[34] The vibrational entropy is given by[33]

$$\frac{S_v}{k_B} = 3\left(\frac{h\nu_D/k_B T}{\exp(h\nu_D/k_B T) - 1} - \ln\left[1 - \exp(-h\nu_D/k_B T)\right]\right). \tag{14}$$

Evaluating Eqs. (12)–(14) at a temperature of 900 °C, which is near the middle of the temperature range considered[10] by Tan and Gösele, the entropy of migration is found to be

$$\Delta S_m/k_B \simeq 3. \tag{15}$$

We will use this as an estimate of the migration entropy for nnh or 2nnh of Ga or As. The masses of Ga and As are very similar so there is very little difference in ΔS_m regardless of which atom hops. Likewise, ΔS_m for nnh and 2nnh are very similar according to the above theory since the difference in ΔS_m is reflected in the hopping volume and the minimum kinetic energy for hopping which are weakly (logarithmically) related to ΔS_m.

We assume[33,35] that the ionization entropy for each ionization process is given by the full entropy of the band gap,

$$\Delta S_i(T) = \Delta S_{cv}(T) = \alpha T(T + 2\beta)/(T + \beta)^2, \tag{16}$$

where $\alpha = 5.405 \times 10^{-4}$ eV/K and $\beta = 204$ K.[36] Using these values for α and β at a temperature of 900 °C gives

$$\Delta S_{cv}/k_B \simeq 6. \tag{17}$$

As a crude approximation,[33] we assume the entropy of formation of Ga or As vacancies in GaAs to be

$$\Delta S_f/k_B \simeq 3. \tag{18}$$

A discussion of some of the issues involved in estimating ΔS_f may be found in Refs. 37–39.

With these estimates of ΔS_m, ΔS_i, and ΔS_f, we attempt to rationalize $\Delta S_a/k_B = 26$ as deduced[10] by Tan and Gösele. Consider the six self-diffusion mechanisms listed in Tables V and VI. Note that mechanisms 1, 2, and 5 do not involve ionization of the diffusing specie. Thus, ΔS_a for these mechanisms is just the sum of ΔS_m and ΔS_f which gives $\Delta S_a/k_B \simeq 6$ for mechanisms 1 and 2 whereas $\Delta S_a/k_B \simeq 9$ for mechanism 5. Thus, mechanisms 1, 2, and 5 are not compatible with the activation entropy deduced[10] by Tan and Gösele.

In contrast, mechanisms 3, 4, and 6 involving nnh explicitly depend upon carrier ionization and capture and thus possess an ΔS_i contribution. The saddle-point configuration for mechanism 3 can be described by the following defect reaction in its most common charge state:

$$V_{\mathrm{As}}^{+} + \mathrm{As}_{\mathrm{Ga}}^{x} \underset{\text{As hop}}{\rightarrow} V_{\mathrm{Ga}}^{-} + 2h^{+}. \tag{19}$$

To first order, the ionization entropy for this defect reaction is $\Delta S_i/k_B = \Delta S_{cv}/k_B \simeq 6$ (i.e., one band-gap entropy). This estimate is in accordance with the assertion[20] of Van Vechten that, for example, $\Delta S_i(V_{\mathrm{As}}^{+}) \simeq \Delta S_i(h^{+})$ and $\Delta S_i(V_{\mathrm{Ga}}^{-}) \simeq \Delta S_i(e^{-})$. Likewise, mechanisms 4 and 6 may be likewise described by

$$V_{\mathrm{Ga}}^{-} + \mathrm{Ga}_{\mathrm{As}}^{2-} \underset{\text{Ga hop}}{\rightarrow} V_{\mathrm{As}}^{+} + 4e^{-}, \tag{20}$$

so that $\Delta S_i/k_B \simeq 6$.

Thus, the saddle-point ionization entropy for mechanisms 3, 4, and 6 is $\Delta S_i/k_B \simeq 6$. In order to estimate the total activation entropy for these mechanisms, we must account for two more contributions to the activation entropy. In the remainder of this discussion, we consider only single-vacancy nnh. The first additional contribution to the activation entropy is the configurational entropy associated with the multiplicity of ways of realizing the saddle-point configuration. The configurational entropy for single-vacancy nnh is $\Delta S_{\mathrm{conf}}/k_B = \ln(2)(12) = 3.2$, where 12 arises from the number of possible sixfold rings accessible to an incident, isolated vacancy, and 2 accounts for the fact that hopping around the sixfold ring can occur in either of two directions. The second additional contribution to the activation entropy is due to the formation entropy associated with the antistructure disorder that occurs along the sixfold ring at the saddle-point configuration. It is difficult to estimate the entropy of this defect complex. This entropy will be mainly due to ionization with a small amount associated with atomic positional disorder. As a very crude estimate, assume each of the five antisites contributes 1 k_B to the activation entropy.

Summing up all of these contributions to the activation entropy results in $\Delta S_a/k_B \simeq 20$ for single-vacancy nnh, mechanisms 3 and 4. Although this value is not in accurate quantitative agreement with $\Delta S_a/k_B$ as deduced[10] by Tan and Gösele, it is an admittedly crude estimate which does illustrate that a large activation entropy is expected for nnh.

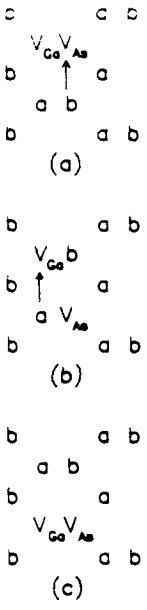


FIG. 3. Divacancy diffusion by second-nearest-neighbor hopping and initiated by an As hop.

are compatible with the conclusion that 2nnh predominates. Note that the activation enthalpy for 2nnh is ~3–4 eV, in agreement with that reported by many researchers, as shown in Table I. As reviewed[9] by Deppe and Holonyak, annealing experiments of QWHs indicate that under As-rich conditions intermixing occurs for n-type materials whereas p-type QWHs remain stable and exhibit no intermixing. This trend is compatible with the energetic assessment as follows. From stoichiometric considerations and from Table IV it is evident that V_{Ga} formation is favored under n-type, As-rich conditions; it is also clear from Table V that Ga 2nnh predominates under these conditions.

Certain other self-diffusion trends are incompatible, however, with the conclusion that 2nnh predominates. For example, as reviewed[9] by Deppe and Holonyak, QWHs exhibit intermixing under As-poor, p-type conditions while

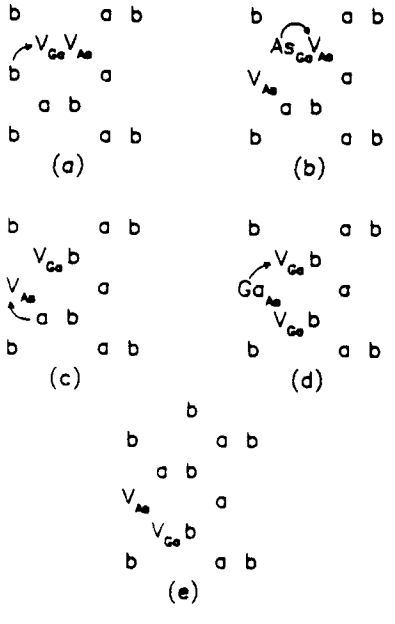


FIG. 4. Divacancy diffusion by nearest-neighbor hopping. Step (d) corresponds to saddle-point A and step (b) corresponds to saddle-point B.

no intermixing occurs when the doping is n type. Under As-poor, p-type conditions V_{As} formation is favored according to Table IV and As 2nnh should predominate according to Table V. As 2nnh is not a viable mechanism for accounting for QWH intermixing under As-poor, p-type conditions, however, since such intermixing is monitored by cation diffusion[9] whereas As 2nnh occurs exclusively on the anion sublattice.

While the atomistic thermodynamic estimates of saddle-point activation enthalpies provide some insight into the nature of QWH intermixing, a detailed understanding of these phenomena requires a consideration of the migration of the dopant species also. For example, Mei *et al.* quote[15] values of the activation energy for Al interdiffusion in AlAs/GaAs superlattices of 4, 2.9, and 1 eV for Si-, Te-, and Zn-doped materials, respectively. The activation energies for Si and Te, both n-type dopants, are comparable with that given in Table V for Ga diffusion by 2nnh and it is possible that differences in the activation energy are at least partially attributable to differences in the Fermi-level position as implied by Table V. However, the 1-eV activation energy for Zn diffusion is definitely incompatible with values given in Table V. For Zn diffusion it appears that the migration of Zn itself by an interstitial-substitutional mechanism, as reviewed in Ref. 9, is the rate-limiting process which determines the impurity-induced intermixing. Note that the BM value for the enthalpy of migration of a Zn atom in GaAs is 0.77 eV for hopping into a hexagonal interstitial site and 1.13 eV for Zn hopping into a body-centered interstitial site.

B. Activation entropy analysis

The self-diffusion energetic analysis up to this point has indicated that 2nnh is the most probable mode of self-diffusion in GaAs. This conclusion is at odds, however, with certain self-diffusion trends and with the self-diffusion analysis[10] of Tan and Gösele who deduce an activation energy of 6 eV for intrinsic material in contrast to our value of 3.97 eV for Ga 2nnh. Note that this 6-eV activation energy does not correspond to any of the ΔH_A's listed in Table V, although it is reasonably close to mechanism 5, divacancy diffusion by 2nnh.

To further investigate the mechanism of self-diffusion in intrinsic GaAs we adopt a different strategy and undertake a thermodynamic analysis of the prefactor D_0. The prefactor may be alternatively expressed[33] in terms of an entropy of activation ΔS_a, which is given as

$$\frac{\Delta S_a}{k_B} = \ln \frac{8D_0}{a^2 f \nu_D}, \tag{10}$$

where a is the lattice parameter (5.6533 Å), f is the correlation factor (1/2), and ν_D is the Debye frequency (7.17×10^{12} s^{-1}) which we use to approximate the attempt frequency. Employing these values and using $D_0 = 2.9 \times 10^8$ cm^2 s^{-1} as deduced[10] by Tan and Gösele in Eq. (10) leads to $\Delta S_a/k_B = 26$, an extraordinarily large value of the entropy of activation. Alternatively, adopting $D_0 = 7 \times 10^{-1}$ cm^2 s^{-1} as reported[4] by Kendall gives

TABLE VI. Summary of thermodynamic expressions used in the self-diffusion energetics analysis. The asterisks denote saddle point charge states which are influenced by association.

Diffusion mechanism	Fermi-level position	Saddle-point constituents and charge state	Expression used to calculate self-diffusion activation energy, ΔH_A
1. Ga migration by 2nnh	E_C	V_{Ga}^-	$\Delta H_f(V_{Ga}^-) + \Delta H_m^{2nn}(Ga)$
	E_I	V_{Ga}^-	$\Delta H_f(V_{Ga}^-) + \Delta H_m^{2nn}(Ga)$
	E_V	V_{Ga}^+	$\Delta H_f(V_{Ga}^+) + \Delta H_m^{2nn}(Ga)$
2. As migration by 2nnh	E_C	V_{As}^x	$\Delta H_f(V_{As}^x) + \Delta H_m^{2nn}(As)$
	E_I	V_{As}^+	$\Delta H_f(V_{As}^+) + \Delta H_m^{2nn}(As)$
	E_V	V_{As}^+	$\Delta H_f(V_{As}^+) + \Delta H_m^{2nn}(As)$
3. Cation vacancy migration by nnh	E_C	$V_{As}^x + 3As_{Ga}^x + 2Ga_{As}^{2-}$	$\Delta H_f(V_{As}^x) + 3\Delta H_f(As_{Ga}^x) + 2\Delta H_f(Ga_{As}^{2-}) + 4E_{coul}^{2nn} + \Delta H_m^{nn}(As)$
	E_I	$V_{As}^+ + 3As_{Ga}^{2+} + 2Ga_{As}^{2-}$*	$\Delta H_f(V_{As}^+) + 3\Delta H_f(As_{Ga}^{2+}) + 2\Delta H_f(Ga_{As}^{2-}) - 12E_{coul}^{nn} + 8E_{coul}^{2nn} - 6E_{coul}^{3nn} + \Delta H_m^{nn}(As)$
	E_V	$V_{As}^+ + 3As_{Ga}^{2+} + 2Ga_{As}^{2-}$*	$\Delta H_f(V_{As}^+) + 3\Delta H_f(As_{Ga}^{2+}) + 2\Delta H_f(Ga_{As}^{2-}) - 12E_{coul}^{nn} + 8E_{coul}^{2nn} - 6E_{coul}^{3nn} + \Delta H_m^{nn}(As)$
4. Anion vacancy migration by nnh	E_C	$V_{Ga}^- + 3Ga_{As}^{2-} + 2As_{Ga}^x$	$\Delta H_f(V_{Ga}^-) + 3\Delta H_f(Ga_{As}^{2-}) + 2\Delta H_f(As_{Ga}^x) + 4E_{coul}^{nn} + 12E_{coul}^{2nn} + 2E_{coul}^{3nn} + \Delta H_m^{nn}(As)$
	E_I	$V_{Ga}^- + 3Ga_{As}^{2-} + 2As_{Ga}^{2+}$*	$\Delta H_f(V_{Ga}^-) + 3\Delta H_f(Ga_{As}^{2-}) + 2\Delta H_f(As_{Ga}^{2+}) - 12E_{coul}^{nn} + 8E_{coul}^{2nn} - 6E_{coul}^{3nn} + \Delta H_m^{nn}(As)$
	E_V	$V_{Ga}^- + 3Ga_{As}^{2-} + 2As_{Ga}^{2+}$*	$\Delta H_f(V_{Ga}^-) + 3\Delta H_f(Ga_{As}^{2-}) + 2\Delta H_f(As_{Ga}^{2+}) - 12E_{coul}^{nn} + 8E_{coul}^{2nn} - 6E_{coul}^{3nn} + \Delta H_m^{nn}(As)$
5. Divacancy migration by 2nnh	E_C	$V_{As}^x + V_{Ga}^-$	$\Delta H_f(V_{As}^x) + \Delta H_f(V_{Ga}^-) + \Delta H_m^{2nn}(Ga)$
	E_I	$V_{As}^+ + V_{Ga}^-$	$\Delta H_f(V_{As}^+) + \Delta H_f(V_{Ga}^-) - E_{coul}^{3nn} + \Delta H_m^{2nn}(Ga)$
	E_V	$V_{As}^+ + V_{Ga}^-$*	$\Delta H_f(V_{As}^+) + \Delta H_f(V_{Ga}^-) - E_{coul}^{3nn} + \Delta H_m^{2nn}(Ga)$
6. Divacancy migration by nnh	E_C	$2V_{As}^x + As_{Ga}^x$(B)	$2\Delta H_f(V_{As}^x) + \Delta H_f(As_{Ga}^x) + \Delta H_m^{nn}(As)$
	E_I	$2V_{Ga}^- + Ga_{As}^{2-}$(A)	$2\Delta H_f(V_{Ga}^-) + \Delta H_f(Ga_{As}^{2-}) + 4E_{coul}^{nn} + E_{coul}^{2nn} + \Delta H_m^{2nn}(Ga)$
	E_V	$2V_{Ga}^x + Ga_{As}^x$(A)	$2\Delta H_f(V_{Ga}^x) + \Delta H_f(Ga_{As}^x) + \Delta H_m^{nn}(Ga)$

most probable diffusion mechanism in terms of an enthalpy of activation analysis may be undertaken. It is evident from Table V that the energetics of self-diffusion strongly depend on the Fermi-level position and thus on the charge state of the defect complexes participating in self-diffusion. In terms of the considerations of the previous section, this is further evidence of the importance of the electronic contribution to the enthalpy of formation of defects.

According to the estimates given in Table V, the most probable self-diffusion mechanism (i.e., the mechanism with the minimum ΔH_A) is seen to always involve 2nnh; Ga diffusion predominates when the Fermi level is near E_c whereas As diffusion predominates when the Fermi level is near E_v. When the Fermi level is near the middle of the band gap, Ga diffusion by 2 nnh is most energetically favorable but As diffusion by 2 nnh has a ΔH_A of nearly the same magnitude and, therefore, cannot be ruled out as a possible self-diffusion mechanism. The clear trend evident from Table V is that self-diffusion in n-type GaAs is predicted to proceed by Ga 2nnh whereas p-type self-diffusion occurs by As 2nnh. An energetic assessment, as given in Table V, indicates that nnh, involving either single vacancies or divacancies, is not an energetically favorable self-diffusion mechanism.

Certain self-diffusion trends observed experimentally

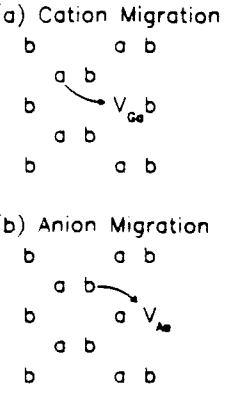


FIG. 1. Vacancy migration by (a) cation and (b) anion second-nearest-neighbor hopping.

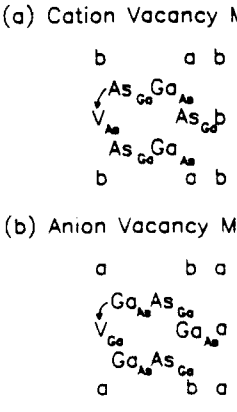


FIG. 2. Saddle-point configuration for (a) cation and (b) anion vacancy migration by nearest-neighbor hopping.

Other assumptions implicit in our atomistic thermodynamic formulation are with regard to the ionization state and energy of isolated and complexed point defects. With regard to the ionization state, we make the simplest possible assumption,[28] motivated by simple chemical intuition, that column-III cation and column-V anion vacancies behave as single acceptors and donors, respectively; likewise, cation-on-anion site and anion-on-cation site antisites are presumed to be double acceptors and donors, respectively. Clearly, considerations such as negative U imply that such assumptions must be made with caution; indeed, other workers have concluded[29] that there are a multiplicity of charge states associated with isolated and complexed point defects in GaAs. Our view is that the chemical intuition assumptions should be employed until experimental analysis forces one to abandon them. The experimental situation regarding the ionization state of isolated point defects in GaAs favors the simple chemical intuition assumptions.[21,29–32]

The isolated point-defect ionization energies for GaAs as listed in Table II are subject to some uncertainty. It will be shown, however, that the main conclusions regarding self-diffusion trends do not sensitively depend on the precise position in energy of these ionization levels.

A related concept that we frequently employ in this analysis in a qualitative manner is the idea of the association of charged defects as formulated[28] by Kröger. If two defects are oppositely charged and occupy nearby lattice sites, the enthalpy of formation of the two-defect complex is reduced in comparison with the sum of the enthalpy of formation of the two isolated defects by a binding energy which is approximately that estimated by the fully screened point-charge Coulombic interaction energy. Additionally, the ionization energies of these two defects are pushed towards their respective band edges by an amount equal to this Coulombic interaction energy. Thus, association of oppositely charged defects results in a perturbation of the ionization energies of the isolated defects which we can estimate semiquantitatively. For example, the saddle-point-charge states denoted by asterisks in Table VI are those expected when the association of oppositely charged defects is considered. If nearby ionized defects possess similar charge states, we can estimate the last-to-ionize defect ionization energy as increased by the extra Coulombic interaction energy due to the neighboring ionized defect.

The final assumption made in this analysis is that the position of the Fermi level is known. The Fermi-level position determines the charge state of the intrinsic point defects present and is itself established from the condition of charge balance at a given temperature. Thus, the only temperature dependence in this formulation is the implicit dependence associated with the Fermi-level position.

III. SELF-DIFFUSION ENERGETICS

A. Activation enthalpy analysis

A summary of the self-diffusion energetics is presented in Table V in which the enthalpy of activation ΔH_A for the saddle-point configuration is estimated for the diffusion mechanism of interest. The enthalpy of activation is assumed to consist of a sum of the enthalpy of formation of the saddle-point-defect complex plus the relevant migration enthalpy as estimated from the ballistic model. A summary of the thermodynamic expressions used in the self-diffusion analysis is given in Table VI. Notice that we consider only vacancy migration modes in our analysis. Although our list of migration modes is not exhaustive, most of the commonly proposed III-V vacancy self-diffusion mechanisms have been considered. Before discussing the conclusions evident from an analysis of Table V, a short description of these self-diffusion mechanisms and their energetics is given.

TABLE V. Summary of self-diffusion energetics.

Diffusion mechanism	ΔH_A (eV) $E_F=E_C$	$E_F=E_I$	$E_F=E_i$
1. Ga diffusion by 2nnh	3.21	3.97	4.72
2. As diffusion by 2nnh	4.90	4.29	3.53
3. Cation migration by nnh	10.40	8.36	6.08
4. Anion migration by nnh	12.94	8.16	10.44
5. Divacancy migration by 2nnh	5.52	5.44	5.44
6. Divacancy migration by nnh[a]	7.21(B)	8.19(A)	8.64(A)

[a]A or B denotes the rate-determining saddle point, as shown in Fig. 4.

Diffusion mechanisms denoted 1 and 2 in Table V are conceptually the simplest modes of vacancy self-diffusion possible in a compound semiconductor since atoms remain on their own sublattice when they undergo second-nearest-neighbor hopping (2nnh) as shown in Fig. 1. The energetics of 2nnh are quite simple, as evident from Table VI, and consist of the sum of the enthalpy of formation of the vacancy in a charge state determined by the Fermi-level position, and the migration enthalpy required for an atom on the same sublattice to accomplish a 2nnh to fill the vacancy.

Diffusion mechanisms 3 and 4 are those proposed[20] by Van Vechten and involve the creation and annihilation of antisite defects around a sixfold ring in order to accomplish vacancy migration without propagating a path of antisite disorder. The saddle-point configurations for cation and anion vacancy nearest-neighbor hopping (nnh) around a sixfold ring are shown in Fig. 2. The energetics are quite complicated and strongly Fermi-level dependent, as evident from Tables V and VI. Note that nnh involves the movement of atoms of both sublattices as the antisite disorder is created and annihilated in the 11-stage process.

We consider two modes of divacancy diffusion which are denoted 5 and 6 in Table V. These diffusion modes are illustrated in Figs. 3 and 4. Note that mechanism 5 is illustrated with the divacancy diffusion initiated by As 2nnh; this type of divacancy diffusion could also be initiated by Ga 2nnh, but the saddle-point configuration is energetically less favorable. Also note that two saddle points, labeled A and B in Fig. 4, are present in mechanism 6, which is due to the fact that both types of atoms hop during this migration mode.

Now that the diffusion mechanisms and their energetics have been briefly summarized, an assessment of the

TABLE IV. Enthalpies of formation of isolated antisite and vacancy defects of various charge states for GaAs. These values were calculated for three Fermi-level positions and by assuming that the antisite ionization energies are those given in Table I. Enthalpy is expressed in units of eV.

E_F position	$\Delta H_f(V_{As}^x)$	$\Delta H_f(V_{As}^+)$	$\Delta H_f(V_{Ga}^x)$	$\Delta H_f(V_{Ga}^-)$
E_c	2.31	2.46	2.31	0.80
E_i	2.31	1.70	2.31	1.56
E_v	2.31	0.94	2.31	2.32

E_F position	$\Delta H_f(As_{Ga}^x)$	$\Delta H_f(As_{Ga}^+)$	$\Delta H_f(As_{Ga}^{+2})$	$\Delta H_f(Ga_{As}^x)$	$\Delta H_f(Ga_{As}^-)$	$\Delta H_f(Ga_{As}^{-2})$
E_C	1.62	2.39	3.39	3.11	1.67	0.35
E_i	1.62	1.63	1.87	3.11	2.43	1.87
E_v	1.62	0.87	0.35	3.11	3.19	3.39

(3) The disorder contribution to the enthalpy of formation of an isolated, neutral antisite defect [i.e., $\Delta H_0(A_B^x)$ and $\Delta H_0(B_A^x)$] is that calculated[20,22,24] from the Phillips–Van Vechten two-band dielectric model. These values are equal to the enthalpy of formation of the neutral antisite defect and are listed in Table IV.

(4) The enthalpy of formation of an isolated, ionized antisite defect [i.e., $\Delta H_f(A_B^-)$, $\Delta H_f(A_B^{2-})$, $\Delta H_f(B_A^+)$, and $\Delta H_f(B_A^{2+})$] is estimated[22] as the sum of the disorder and electronic contributions. The electronic contribution is related to the ionization energies of the isolated antisite defect which are listed in Table II. The explicit atomistic thermodynamic equations used to find the total enthalpy of formation for an isolated antisite defect are collected in Table III. Note that only the Fermi-level position is required to find the electronic contribution of the enthalpy of formation of the full ionized antisite defect, whereas the antisite ionization energies are required to determine the electronic contribution for the neutral or partially ionized antisite defect.

(5) The binding energy of an ionized, intrinsic defect is found[25] using the fully screened point charge Coulombic energy as an estimate of the electronic contribution to the binding energy,

$$\Delta E_{coul}=\frac{Z_1 Z_2 q^2}{4\pi\epsilon r_{12}}, \tag{7}$$

where Z_1 and Z_2 are the charges of the respective point defects being considered, r_{12} is the equilibrium separation of the two point defects assuming they occupy ideal lattice sites, and ϵ is the dielectric constant. Assuming $\epsilon = 13.1$ for GaAs, $E_{coul}^{nn} = 0.45$ eV, $E_{coul}^{2nn} = 0.39$ eV, $E_{coul}^{3nn} = 0.23$ eV, and $E_{coul}^{4nn} = 0.19$ eV for first through fourth-nearest-neighbor defect separation, respectively.

(6) The binding enthalpy of a neutral divacancy $\Delta H_b(V_{Ga}V_{As}^x)$, is 0.98 eV, as estimated[20,25] from the MCM model. We assume the divacancy to have a neutral charge state, $V_{As}^+V_{Ga}^-$, for all Fermi-level positions considered; this assumption is motivated by the association of charged species, as discussed below. The enthalpy of formation of the neutral divacancy is found to be 1.83 eV and is calculated according to the following expression:

$$\Delta H_f(V_{As}^+V_{Ga}^-)=\Delta H_f(V_{As}^+)+\Delta H_f(V_{Ga}^-) - \Delta H_b(V_{As}V_{Ga}^x)-E_{coul}^{nn}. \tag{8}$$

(7) The enthalpy of migration, ΔH_m, is estimated[20,26] via the ballistic model (BM) of atomic migration. The migration enthalpies for GaAs from the BM are as follows: $\Delta H_m^{nn}(Ga) = 0.91$ eV, $\Delta H_m^{2nn}(Ga) = 2.41$ eV, $\Delta H_m^{nn}(As) = 0.97$ eV, and $\Delta H_m^{2nn}(As) = 2.59$ eV, where the superscript indicates nearest- or second-nearest-neighbor hopping and the parentheses indicate which atom hops.

It is appropriate to make several comments regarding the assumptions of the present analysis. Perhaps the most difficult aspect of defect energetics calculations is to accurately account for the binding enthalpy of a defect complex. Binding enthalpy is the difference in enthalpy between the isolated point defects which comprise a defect complex and the enthalpy of the defect complex itself. For a stable defect complex, the enthalpy cost of the defect complex should be less than that of the isolated defect complex constituents. In general, the binding enthalpy can be expressed as a sum of various contributions,

$$\Delta H_b=\Delta H_{electronic}+\Delta H_{disorder}+\Delta H_{elastic}, \tag{9}$$

where the subscripts describe the nature of the binding energy contribution. The electronic contribution to the binding energy may be estimated from total energy calculations in which exchange and correlation are included.[27] The disorder term accounts for local fluctuations in the band gap due to antisite disorder;[20,22,24] when defect complexes involving antisite defects form, the disorder contribution should differ from that of isolated antisites. The elastic contribution arises from mechanical, strain energy due to the formation of a defect complex. Since Ga and As atoms are almost identical in size, the elastic contribution in GaAs should be relatively small. All three of these contributions will lead to a certain amount of relaxation or reconstruction in which the defect complex constituents, as well as the neighboring atoms surrounding the complex, occupy positions different from their ideal lattice sites.

Having discussed the subtleties of properly accounting for the binding enthalpy, we reiterate that our treatment of the binding enthalpy involves accounting for the Coulombic contribution to the electronic binding energy using the fully screened, unrelaxed, point-charge approximation and the MCM treatment to account for the divacancy binding energy.

also that a large activation energy is concomitant with an exceedingly large prefactor.

The purpose of the work discussed herein is to estimate the energetics of GaAs self-diffusion in order to elucidate possible self-diffusion mechanisms. Self-diffusion modes involving vacancies are of primary consideration, although interstitials in the form of Frenkel pairs are invoked in the discussion section. First, the activation enthalpies for various modes of self-diffusion are estimated. The minimum saddle-point activation enthalpy for vacancy self-diffusion mechanisms occurs for 2nnh with a value of ~3–4 eV, depending upon the location of the Fermi level. Second, the entropy of activation for several modes of self-diffusion is estimated. It is concluded that the magnitude of the entropy of activation for 2nnh is consistent with normal self-diffusion prefactors of magnitude $D_0 \simeq 10^{-5}$–10^{-1} $cm^2\,s^{-1}$. It is also shown that self-diffusion mechanisms involving nearest-neighbor hopping (nnh) possess very large activation entropies and energies. Although the energetic analysis indicates nnh to be characterized by abnormally large activation entropies and enthalpies, the deduced values are not in quantitative agreement with the self-diffusion parameters reported[10] by Tan and Gösele. Furthermore, it is difficult to understand why, if vacancies are available, self-diffusion should occur by nnh instead of 2nnh since 2nnh is much more energetically favorable. Therefore, according to our self-diffusion energetic estimates we conclude that *if vacancies are available* in equilibrium concentrations, self-diffusion should proceed by 2nnh with self-diffusion characterized by $D_0 \simeq 10^{-5}$–10^{-1} $cm^2\,s^{-1}$ and $E_a \simeq 3$–4 eV. On the other hand, if surface-vacancy generation is precluded by the fact that diffusion occurs in the bulk, remote from surfaces and dislocations, we suggest that the rate-limiting step of self-diffusion is Frenkel pair generation which is characterized by $D_0 \simeq 10^7$–10^8 $cm^2\,s^{-1}$ and $E_a \simeq 6$ eV for cation self-diffusion.

II. INTRINSIC POINT-DEFECT ENERGETICS

Atomistic thermodynamics is the application of macroscopic thermodynamic principles to the study of materials on the atomic scale. The approach adopted in this paper relies heavily on the work of Van Vechten, as reviewed in Ref. 20. The underlying premise of the present work is that microscopic identification of the point-defect complex or defect mechanism can be accomplished by calculating the enthalpy of formation ΔH_f of various possible defects or mechanisms, the one with the smallest ΔH_f being the most likely to occur. Although at constant temperature and pressure the thermodynamic potential of relevance is the Gibbs free energy ΔG_f, it is common practice, and usually sufficiently accurate, to neglect the entropy contribution to ΔG_f and to thus approximate ΔG_f by ΔH_f. We will implicitly assume this to be the case in the following treatment.

The following assumptions are made in order to make the atomistic thermodynamic analysis tractable to the problem at hand:

TABLE II. Estimated ionization energies at $T = 0$ K of isolated antisite and vacancy defects in GaAs.[a] Energies are with respect to the conduction-band minimum for donors and the valence-band maximum for acceptors.

Defect	Ionization level	Charge state	Ionization energy (eV)
Ga_{As}	A1	−/0	0.078
	A2	−/−−	0.20
As_{Ga}	D1	0/+	0.77
	D2	+/++	1.00
V_{Ga}	A	−/0	0.01
V_{As}	D	0/+	0.15

[a]See Ref. 21.

(1) The enthalpy of formation of isolated, neutral vacancies in GaAs [i.e., $\Delta H_f(V^x_{Ga}) = 2.31$ eV and $\Delta H_f(V^x_{As}) = 2.31$ eV] are estimated from the microscopic cavity model (MCM).[20,22]

(2) The enthalpy of formation of an isolated, ionized vacancy in GaAs [i.e, $\Delta H_f(V^-_{Ga})$ and $\Delta H_f(V^+_{As})$] is estimated[20] as the sum of the enthalpy of formation of a neutral vacancy and the electronic contribution; the electronic contribution is related to the ionization energy of the vacancy. The assumed ionization energies for isolated vacancy and antisite defects are listed in Table II. The explicit atomistic thermodynamic equations used to calculate the total enthalpy of formation of an isolated, ionized vacancy are listed in Table III. Note that the position of the Fermi level, as well as the ionization energy of the defect, must be known in order to establish the electronic contribution to the enthalpy of formation. The enthalpies of formation of neutral or ionized, isolated vacancies in GaAs are listed in Table IV for three Fermi-level positions, E_C, E_I, E_V, corresponding to the conduction-band minimum, the intrinsic level, and the valence-band maximum, respectively.

TABLE III. Summary of the thermodynamic expressions used to calculate the enthalpies of formation of isolated antisite and vacancy defects in GaAs.[a,b] ΔH_0 denotes the disorder contribution to the enthalpy of formation.

Defect	Charge state	Expression used to calculate ΔH_f
Ga_{As}	0	$\Delta H_0(Ga^x_{As}) + (E_c - E_{A1}) + (E_c - E_{A2})$
	−	$H_0(Ga^x_{As}) + (E_c - E_F) + (E_c - E_{A2})$
	−−	$\Delta H_0(Ga^x_{As}) + 2(E_c - E_F)$
As_{Ga}	0	$\Delta H_0(As^x_{Ga}) + (E_{D1} - E_v) + (E_{D2} - E_v)$
	+	$\Delta H_0(As^x_{Ga}) + (E_F - E_v) + (E_{D2} - E_v)$
	++	$\Delta H_0(As^x_{Ga}) + 2(E_F - E_v)$
V_{Ga}	0	$\Delta H_f(V^x_{Ga})$
	−	$\Delta H_f(V^x_{Ga}) + (E_A - E_F)$
V_{As}	0	$\Delta H_f(V^x_{As})$
	+	$\Delta H_f(V^x_{As}) + (E_F - E_D)$

[a]See Ref. 20.
[b]See Ref. 23.

TABLE I. Summary of GaAs self-diffusion and superlattice interdiffusion data.

Gallium		Arsenic			
D_0 (cm^2/s)	E_A (eV)	D_0 (cm^2/s)	E_A (eV)	Comments	Reference
10^7	5.6	4×10^{21}	10.2	Self-diffusion	1
	3			Lattice-constant measurements after annealing	2
2.1×10^{-3}	2.1	7×10^{3}	4	Electrical properties measured after annealing	3
7×10^{-1}	3.2			Self-diffusion	4
4.6×10^{-5} ($\pm 16\times10^{-5}$)	2.6±0.5	5.5×10^{-4} ($\pm 2.4\times10^{-4}$)	3±0.04	Self-diffusion	6–8
2.9×10^{8}	6			Self-diffusion/interdiffusion under intrinsic conditions	10–15
	4			Si-induced superlattice interdiffusion	15
	2.9			Te-induced superlattice interdiffusion	16
	4.7			Strain-free superlattice interdiffusion	18

As reviewed by Deppe and Holonyak, these considerations have led certain researchers to conclude that intermixing under As-rich conditions is mediated by column-III vacancies while As-poor intermixing trends are attributed to column III self-interstitials.

The IILD trends as a function of As overpressure and doping discussed previously should also apply to GaAs self-diffusion. The close interrelationship between GaAs self-diffusion and IILD is perhaps best demonstrated by the work[10] of Tan and Gösele, who found that all of the available Ga self-diffusion in GaAs data as well as Al-Ga interdiffusion data under intrinsic carrier conditions could be accurately fit to

$$D_{Ga}(n_i)=2.9\times10^{8}\exp(-6\ \text{eV}/k_BT)\text{cm}^2\,\text{s}^{-1}. \quad (5)$$

Note that this relation includes the self-diffusion data[1,7] of both Goldstein and Palfrey *et al.* Since the Al-Ga interdiffusion of various investigators[11–15] correlates well with Ga self-diffusion in GaAs, it would appear that Al-Ga interdiffusion is rate limited by the diffusion of Ga. Furthermore, Tan and Gösele conclude that Ga self-diffusion is controlled by the position of the Fermi level and is governed by a triply negatively charged point-defect specie, which they attribute to a triply negatively charged Ga vacancy. Their contention that Ga self-diffusion involves a triply negatively charged defect is based on an assessment of the experimental data[15] of Mei *et al.* in which they find that for Si doping, the ratio of the extrinsic to intrinsic diffusivity scales as the cube of the ratio of the carrier concentration to the intrinsic concentration, i.e.,

$$D_{Ga}(n)/D_{Ga}(n_i)=(n/n_i)^3. \quad (6)$$

This cubic power dependence implies a triply negatively ionized defect.

Mei *et al.* report[15] a single activation energy of ~4 eV for silicon-induced intermixing of AlAs-GaAs superlattices. In further work, Mei *et al.* found[16] that Te-induced AlAs-GaAs superlattice intermixing is characterized by an activation energy of 2.9 eV and that the diffusion coefficient scales linearly with the Te carrier concentration instead of to the third power, as found for Si intermixing. Tan and Gösele have reinterpreted[17] the Te-induced intermixing data of Mei *et al.* assuming that not all of the Te incorporated into the superlattice is electrically active, and have argued that this data may be interpreted to have a cubic dependence on the carrier concentration. Regardless of which interpretation is correct, Mei *et al.*'s data is convincing evidence that superlattice interdiffusion depends on the nature of the dopant employed.

Guido *et al.* have demonstrated[18] the importance of encapsulation during IILD of AlGaAs-GaAs superlattices. They found that the interdiffusion coefficient was approximately one order of magnitude smaller for a Si_3N_4 encapsulant compared to a SiO_2 encapsulant or an As overpressure. This trend is attributed to the fact that Si_3N_4 seals the surface with respect to Ga outdiffusion and is, therefore, not a source of vacancies. In contrast, SiO_2 is porous to Ga outdiffusion and is thus an effective source of vacancies. Likewise, an As overpressure will increase the tendency to form column-III vacancies. Thus, this work provides evidence that Al-Ga interdiffusion is a vacancy-mediated process. Guido *et al.* also report activation energies of ~3.42 and ~3.61 eV for Si_3N_4 and SiO_2, respectively, and ~4.75 eV for an As overpressure. Ralston *et al.* have reported[19] an activation energy of 3.82 eV for SiO_2 encapsulated interdiffusion. Guido *et al.* attribute the smaller activation energy for SiO_2 or Si_3N_4 to strain induced by the encapsulant. Thus, they conclude that the activation energy for Al-Ga interdiffusion is ~4.7 eV under strain-free conditions.

A summary of GaAs diffusion and superlattice interdiffusion data is given in Table I. Note the lack of agreement between various researchers. Although it is difficult to draw any firm conclusions from the existent data, it would appear that researchers deduce an activation energy of either ~3–4 eV or of about 6 eV for Ga diffusion. Note

Energetics of self-diffusion in GaAs

J. F. Wager
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

(Received 9 July 1990; accepted for publication 10 December 1990)

Atomistic thermodynamic calculations are performed in order to examine the energetics of self-diffusion in GaAs. An energetic assessment of the activation enthalpy of the saddle-point configuration of various modes of vacancy self-diffusion indicates second-nearest-neighbor hopping to be the energetically most favorable mechanism if vacancies are available in equilibrium concentrations. An assessment of the activation entropy indicates that normal diffusion prefactors of magnitude $D_0 \simeq 10^{-5}$–10^{-1} s, $cm^2\,s^{-1}$ are consistent with vacancy self-diffusion by second-nearest-neighbor hopping. It is proposed that self-diffusion experiments characterized by prefactors and activation energies of large magnitude, e.g., $D_0 \simeq 10^7$–10^8 $cm^2\,s^{-1}$ and $E_a \simeq 6$ eV, involve processes in which surface vacancy generation is inhibited and self-diffusion is mediated by Frenkel pair generation.

I. INTRODUCTION

The first study[1] of GaAs self-diffusion was performed by Goldstein, using radioactive tracer analysis, who reported the diffusivity of Ga and As to be

$$D_{Ga}=1\times10^{7}\exp(-5.6\ \text{eV}/k_BT)\text{cm}^2\,\text{s}^{-1}, \quad (1)$$

$$D_{As}=4\times10^{21}\exp(-10.2\ \text{eV}/k_BT)\text{cm}^2\,\text{s}^{-1}. \quad (2)$$

Goldstein comments that these activation energies and prefactors are among the largest in the existent literature. Potts and Pearson[2] estimated an activation enthalpy of 3 eV for As vacancy migration in GaAs from lattice constant annealing experiments and note that their estimate is at odds with Goldstein's reported value of 10.2 eV. In later work, Chiang and Pearson reported[3] activation energies of 2.1 and 4.0 eV for gallium and arsenic vacancy migration, respectively.

Kendall questions[4] the accuracy of Goldstein's self-diffusion coefficients, pointing out that self-diffusion in GaAs proceeds at an exceedingly slow rate and that there are formidable experimental difficulties which preclude accurate assessment of these quantities. He also asserts that Goldstein's estimates are too low because of problems associated with the incongruent evaporation of As from the GaAs surface. From a limited number of isolated data points obtained from radioactive isotope measurements, Kendall estimates an As self-diffusion activation energy of 3.2 eV.

The most recent experimental work on GaAs self-diffusion was reported[6,7] by Palfrey *et al.* and reviewed[8] by Willoughby. Radioactive tracer experiments were performed at a constant As pressure [$P(As_2) = 0.75$ atm] over a temperature range of 1025–1100 °C for Ga self-diffusion and 1000–1075 °C for As self-diffusion with the resulting diffusion coefficients as follows:

$$D_{Ga}=(4\times10^{-5}\pm16\times10^{-5})\exp(-2.6\pm0.5\ \text{eV}/k_BT)\text{cm}^2\,\text{s}^{-1}, \quad (3)$$

$$D_{As}=(5.5\times10^{-4}\pm2.4\times10^{-4})\exp(-3.0\pm0.04\ \text{eV}/K_BT)\text{cm}^2\,\text{s}^{-1}. \quad (4)$$

For Ga self-diffusion, Palfrey *et al.* conclude[6] that there is a single mechanism operative rather than a complex one but that the diffusion mechanism cannot be concluded definitively from their data. They note that the very slow rate of self-diffusion and high activation energy is inconsistent with a simple interstitial mechanism and that the observed magnitudes of these quantities can be interpreted as evidence for some type of vacancy mechanism. For As self-diffusion, Palfrey *et al.* note that an increase in the As overpressure to $P(As_2) = 3.0$ atm results in a decrease in the As diffusivity at a constant temperature. They note that this decrease in the diffusivity with increasing As pressure is inconsistent with interstitial diffusion, As migration via As_{Ga} antisite defects moving exclusively on the Ga sublattice, and with diffusion via As_{Ga}-V_{As} pairs. In contrast, such an As pressure dependence is consistent with As second-nearest-neighbor hopping (2nnh).

New insights into the nature of GaAs self-diffusion have been obtained recently from impurity-induced layer disordering (IIID) studies of AlGaAs/GaAs quantum-well heterostructures (QWHs) and superlattices (SLs), as reviewed[9] by Deppe and Holonyak. For example, annealing experiments of QWHs indicate that under As-rich conditions intermixing occurs, as monitored by cation migration, for *n*-type materials whereas *p*-type QWHs remain stable and exhibit no intermixing. In contrast, under As-poor conditions, *p*-type QWHs undergo intermixing but no intermixing occurs when the doping is *n* type. Cation migration, and hence intermixing, of QWHs depends upon the As vapor pressure because diffusion is mediated by native defects whose concentration is a function of the As pressure. As-rich conditions favor the formation of As antisites, column-III vacancies, or As interstitials whereas As-poor conditions favor the formation of column-III antisites, As vacancies, or column-III interstitials. Additionally, as discussed further later in this paper, *n*-type doping favors the formation of acceptor defects whereas *p*-type doping makes donor defects energetically more favorable.

 0021-8979/91/053022-10$03.00

BOCQUET, J. L., BRÉBEL, G., and LIMOGE, Y., 1983, *Physical Metallurgy*, third edition, edited by R. W. Cahn and P. Haasen (Amsterdam: Elsevier), chap. 8.

DOBSON, T. W., WAGER, J. F., and VAN VECHTEN, J. A., 1989, *Phys. Rev.* B, **12**, 2962.

LEE, J. F., SEARS, F. W., and TURCOTTE, D. L., 1973, *Statistical Thermodynamics* (Reading, Massachusetts: Addison-Wesley), chaps. 9, 12.

VAN VECHTEN, J. A., 1975, *Phys. Rev.* B, **12**, 1247.

VAROTSOS, P. A., and ALEXOPOULOS, K. D., 1986, *Thermodynamics of Point Defects and Their Relation with Bulk Properties* (Amsterdam: North-Holland).

Taking the partial derivative of eqn. (B 2) with respect to T at constant V and multiplying by $k_B T^2$ gives

$$k_B T^2[\partial(\ln Z_s)/\partial T]_V = \tfrac{7}{4}k_B T + \mathrm{KE_m}. \tag{B 3}$$

Taking the partial derivative of eqn. (B 2) with respect to V at constant T and multiplying by $k_B TV$ gives

$$k_B TV[\partial(\ln Z_s)/\partial V]_T = k_B T. \tag{B 4}$$

Substitution of eqns. (B 3) and (B 4) into eqn. (B 1) yields

$$H_s = \tfrac{11}{4}k_B T + \mathrm{KE_m}. \tag{B 5}$$

Now evaluate H_i by finding the natural logarithm of Z_i as given by eqn. (26):

$$\ln Z_i = -3\{h\nu_D/k_B T + \ln[1 - \exp(-h\nu_D/k_B T)]\}. \tag{B 6}$$

Taking the partial derivative of eqn. (B 6) with respect to T at constant V and multiplying by $k_B T^2$ gives

$$k_B T^2[\partial(\ln Z_i)\partial T]_V = 3h\nu_D\{\tfrac{1}{2} + 1/[\exp(h\nu_D/k_B T) - 1]\}. \tag{B 7}$$

The partial derivative of Z_i with respect to V is zero so that

$$H_i = 3h\nu_D\{\tfrac{1}{2} + 1/[\exp(h\nu_D/k_B T) - 1]\}. \tag{B 8}$$

From eqns. (B 5) and (B 8) we obtain the desired result:

$$\Delta H_m = H_s - H_i = \mathrm{KE_m} + \tfrac{11}{4}k_B T - 3h\nu_D\{\tfrac{1}{2} + 1/[\exp(h\nu_D/k_B T) - 1]\}. \tag{B 9}$$

APPENDIX C

Evaluation of ΔS_m

The standard statistical thermodynamic expression relating the partition function to the entropy is given by

$$S = k_B\{\ln Z + T[\partial(\ln Z)/\partial T]_V\}. \tag{C 1}$$

Multiplying eqn. (B 2) by k_B, dividing eqn. (B 3) by T, and adding these together give

$$S_s = \tfrac{1}{4}k_B + k_B \ln\{[1/4^{5/4}(2\pi)^{1/2}](k_B T/\mathrm{KE_m})^{1/4}[(2\pi mek_B T)^{3/2}V/h^3]\}. \tag{C 2}$$

Note that $e = 2{\cdot}718$ and this term enters into the second set of square brackets from an additional $\frac{3}{2}k_B$ term that was partitioned from the $\frac{7}{4}k_B$ term in eqn. (B 3) and rewritten as $\frac{3}{2}k_B = k_B \ln(e^{3/2})$. Multiplying eqn. (B 6) by k_B, dividing eqn. (B 7) by T, and adding these together yields

$$S_i = 3k_B\{(h\nu_D/k_B T)/[\exp(h\nu_D/k_B T) - 1] - \ln[1 - \exp(-h_D/k_B T)]\}. \tag{C 3}$$

We obtain ΔS_m from eqns. (C 2) and (C 3):

$$\Delta S_m = S_s - S_i = \tfrac{1}{4}k_B + k_B \ln\{[1/4^{5/4}(2\pi)^{1/2}](k_B T/\mathrm{KE_m})^{1/4}[(2\pi mek_B T)^{3/2}V/h^3]\}$$
$$- 3k_B\{(h\nu_D/k_B T)/[\exp(h\nu_D/k_B T) - 1] - \ln[1 - \exp(-h\nu_D/k_B T)]\}. \tag{C 4}$$

References

Beatty, J. W., and Scamehorn, R. G., 1977, *Reaction Kinetics* (Washington, DC: University Press of America), pp. 87–105.

Employ a change in variable,

$$\alpha^2 = p_x^2/2mk_BT \tag{A 2}$$

which leads to

$$\int_{p_{min}}^{\infty} \exp\left(-\frac{p_x^2}{2mk_BT}\right) dp_x = (2mk_BT)^{1/2} \int_{\alpha_{min}}^{\infty} \exp(-\alpha^2)\, d\alpha. \tag{A 3}$$

Break up the integral on the right-hand side into two integrals:

$$\int_{\alpha_{min}}^{\infty} \exp(-\alpha^2)\, d\alpha = \int_0^{\infty} \exp(-\alpha^2)\, d\alpha - \int_0^{\alpha_{min}} \exp(-\alpha^2)\, d\alpha. \tag{A 4}$$

Note that

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-x^2)\, dx \tag{A 5}$$

and

$$\text{erf}(\infty) = 1 \tag{A 6}$$

so that eqn. (A 4) can be simplified to

$$\int_{\alpha_{min}}^{\infty} \exp(-\alpha^2)\, d\alpha = \frac{\sqrt{\pi}}{2} - \frac{\sqrt{\pi}}{2} \text{erf}(\alpha_{min}). \tag{A 7}$$

Note that the error function can be expanded (Lee *et al.* 1973) as

$$\text{erf}(x) = 1 - \{[\exp(-x^2)]/(\pi x)^{1/2}\}[1 - 1/2x^2 + 1 \times 3/(2x^2)^2 - 1 \times 3 \times 5/(2x^2)^3 + \ldots]. \tag{A 8}$$

Keeping only the first two terms in the expansion (note that this approximation is valid as long as $KE_m \gg k_BT$) and substituting back into eqn. (A 7) give

$$\int_{\alpha_{min}}^{\infty} \exp(-\alpha^2)\, d\alpha = \exp(-\alpha_{min}^2)/(4\alpha_{min})^{1/2}. \tag{A 9}$$

Substituting eqn (A 9) into eqn. (A 3) using eqn (A 2), and recognizing that $KE_m = p_x^2/2m$ finally yields the desired result

$$\int_{p_{min}}^{\infty} \exp\left(-\frac{p_x^2}{2mk_BT}\right) dp_x = \left(\frac{m^2(k_BT)^3}{4KE_m}\right)^{1/4} \exp\left(-\frac{KE_m}{k_BT}\right). \tag{A 10}$$

APPENDIX B

Evaluation of ΔH_m

We begin with the standard statistical thermodynamic relationship

$$H = k_BT^2[\partial(\ln Z)/\partial T]_V + k_BTV[\partial(\ln Z)/\partial V]_T. \tag{B 1}$$

First evaluate H_s by finding the natural logarithm of Z_s as given by eqn. (25):

$$\ln Z_s = \ln[1/4^{5/4}(2\pi)^{1/2}] + \tfrac{1}{4}\ln(k_BT/KE_m) + \tfrac{3}{2}\ln(2\pi mk_BT) + \ln(V/h^3) - KE_m/k_BT. \tag{B 2}$$

interpreted as evidence for divacancy migration in addition to that of single vacancies (for example Bocquet *et al.* (1983)). The present theory predicts a temperature dependence of opposite sense to that required to account for Arrhenius plots that curve upwards. It should be recognized, however, that the temperature dependences of the pre-factor and activation enthalpy are given by a superposition of the formation as well as the migration terms. Thus it is possible that an increasing formation contribution could overcompensate for the decreasing migration component. It is more likely, however, that the decreases in ΔH_m and ΔS_m with increasing temperature are a consequence of employing a rather crude model to estimate the vibrational entropy in the initial configuration prior to the hop (i.e. the entropy of the initial configuration is estimated as that of an atom with four nearest neighbours within an Einstein solid which is vibrating at the Debye frequency). Note from table 2 that S_s, the translational contribution to ΔS_m evaluated at the saddle point, increases with increasing temperature as does S_i, the vibrational contribution to ΔS_m evaluated at the initial configuration. ΔS_m decreases with increasing temperature only because S_i increases with temperature faster than does S_s. A more realistic treatment of S_i might result in a temperature dependence more compatible with the hypothesis that curved Arrhenius plots arise from the temperature dependence of the prefactor and activation enthalpy.

§4. Conclusions

A statistical thermodynamic derivation of the enthalpy ΔH_m of migration and entropy ΔS_m of migration by nearest-neighbour hopping and in accordance with the BM of vacancy migration is presented. Although the general forms of ΔH_m and ΔS_m are similar to those given previously (Van Vechten 1975, Dobson *et al.* 1989), there are correction terms to both of these quantities which arise from the statistical thermodynamic derivation. Except at very high temperatures or for a material with an exceedingly large Debye temperature, the corrections to ΔH_m are of negligible importance. In contrast, the corrections to ΔS_m are significant; agreement between experimental and calculated values of ΔS_m for various elemental metals which are known to diffuse via nearest-neighbour vacancy migration can only be obtained when these corrections to ΔS_m are employed. The present theory predicts ΔH_m and ΔS_m to decrease with increasing temperature; this trend is attributed to the rather crude model employed to estimate the vibrational entropy of the initial configuration.

Acknowledgements

The author wishes to thank J. A. Van Vechten and T. W. Dobson for useful discussions. This work was supported by the U.S. Air Force Office of Scientific Research under Contract AFOSR 89-0309.

APPENDIX A

Evaluation of the p_x integral

It is desired to evaluate the following integral:

$$\int_{p_{min}}^{\infty} \exp\left(-\frac{p_x^2}{2mk_BT}\right) dp_x. \tag{A 1}$$

using experimental values of these quantities as compiled by Varotsos and Alexopoulos (see Varotsos and Alexopoulos (1986) for a detailed review of the experimental data). The rather wide range of experimental values for $\Delta S_m/k_B$ is a consequence of differences in sample preparation, experimental technique, and method of data analysis. It is clear from an analysis of table 1 that, with the exception of Pb, the present method of calculating ΔS_m via eqn. (28) gives much better agreement with experiment, and also with the cBΩ theory of Varotsos and Alexopoulos, than does the previous method given by eqns. (4)–(6). It should be noted that the large values of $\Delta S_m/k_B$ for Pb arise because there is only one report for the entropy of formation and this value seems anomalously low and has a large degree of uncertainty (i.e. $\Delta S_f/k_B = 0{\cdot}7 \pm 2{\cdot}0$). Thus it is evident from table 1 that ΔS_m is overestimated when computed via eqns. (4)–(6).

According to the present theory, ΔH_m and in particular ΔS_m are temperature-dependent quantities as evident from eqns. (27) and (28). The temperature dependence of these quantities is indicated in table 2 for three temperatures: T_m, $0{\cdot}75T_m$ and $0{\cdot}5T_m$, where T_m is the melting temperature. The translational entropy S_s of the saddle-point configuration and the vibrational entropy S_i of the initial state are also included in table 2. Note from table 2 that, for the metals considered, the temperature dependence of ΔH_m is indeed very small such that eqn. (29) is well justified.

It has often been asserted that curved Arrhenius plots arise from the temperature dependence of the enthalpies and entropies of formation and migration (for example Varotsos and Alexopoulos 1986). (Alternatively, curved Arrhenius plots have also been

Table 2. Entropies of translation, ionization and migration and enthalpy of migration for various elements evaluated at T_m, $0{\cdot}75T_m$ and $0{\cdot}5T_m$, where T_m is the melting temperature.

Element	Temperature (K)	S_s/k_B	S_i/k_B	$\Delta S_m/k_B$	ΔH_m (eV)
Al	933	7·27	5·36	1·91	0·716
Al	700	6·76	4·52	2·24	0·719
Al	467	6·05	3·37	2·68	0·722
Cu	1110	8·62	6·54	2·08	0·864
Cu	833	8·12	5·68	2·44	0·869
Cu	555	7·41	4·49	2·92	0·873
Ag	1234	10·01	8·12	1·89	0·803
Ag	926	9·51	7·25	2·26	0·809
Ag	617	8·80	6·04	2·76	0·815
Pb	601	10·28	8·23	2·05	0·497
Pb	450	9·7	7·38	2·39	0·500
Pb	300	9·05	6·16	2·89	0·503
Na	371	6·57	5·58	0·99	0·101
Na	278	6·07	4·73	1·34	0·102
Na	158	5·08	3·56	1·52	0·103
Li	454	4·39	3·90	0·49	0·095
Li	340	3·89	3·09	0·80	0·095
Li	227	3·18	2·02	1·16	0·094
K	1204	10·23	10·75	−0·52	0·074
K	903	9·72	9·89	−0·17	0·080
K	602	9·02	8·67	0·35	0·087
W	3683	12·30	9·66	2·64	3·010
W	2762	11·80	8·80	3·00	3·119
W	1842	11·09	7·59	3·50	3·138

temperature derivatives of the normal translational partition function enclosed within the curly brackets in eqn. (25), a $\frac{1}{4}k_BT$ contribution from the temperature derivative of the $(k_BT/KE_m)^{1/4}$ term in eqn. (25) and a k_BT contribution from the volume derivative of the normal translational partition function enclosed within the curly brackets in eqn. (25). The second correction term is recognized as the vibrational energy of an Einstein solid evaluated at the Debye frequency; this term arises from the temperature derivative of the vibrational partition function Z_i. Except at a high temperature or when the Debye temperature is very large, the correction terms will be quite small and, to first order,

$$\Delta H_m \approx KE_m. \tag{29}$$

The migrational entropy ΔS_m is also the same as that proposed previously (eqns. (4)–(6)) but with a correction term of $\frac{1}{4}k_B$ and an additional logarithmic correction. In order to assess the importance of these correction terms to ΔS_m which arise from the statistical thermodynamic derivation, a comparison between ΔS_m values for elemental metals calculated via the previous method (Dobson *et al.* 1989) and by the present method is indicated in table 1. Also included in table 1 are ΔS_m values calculated according to the cBΩ model (Varotsos and Alexopoulos 1986); these values are in good agreement with values deduced experimentally although it should be emphasized that there is often a wide range of variation in values of ΔS_m deduced experimentally. The elemental metals listed in table 1 were chosen because of the availability of experimental data for ΔS_m and since it is well established for these metals that, at temperatures well below the melting temperature, self-diffusion occurs by nearest-neighbour hopping of single vacancies (Bocquet, Brébec and Limoge 1983).

The importance of the correction terms which arise from the statistical thermodynamic derivation is clearly established in table 1 in which the present theory is compared with the previous theory of Dobson *et al.*, to the cBΩ theory of Varotsos and Alexopoulos, and to experimental values of ΔS_m as compiled by Varotsos and Alexopoulos. Except for Cu, the experimental values for $\Delta S_m/k_B$ indicated in table 1 are obtained by subtracting the entropies of formation from the entropies of activation

Table 1. Migration entropies of various elemental metals calculated according to the theory of Dobson *et al.* (1989), Wager (this work), and Varotsos and Alexopoulos (1986). The experimental values are taken from Varotsos and Alexopoulos.

	$\Delta S_m/k_B$			
Element	Dobson *et al.*†	This work†	(Varotsos and Alexopoulos)	Experimental (Varotsos and Alexopoulos)
Al	5·27	2·24	2·99	2·1–4·1
Cu	5·47	2·44	1·72	1·65, 2·25
Ag	5·25	2·26	≈0–2·3	≈2–3·5
Pb	5·43	2·39	2·37	≈2·5–5·9
Na	3·74	1·02	1·32	≈1·4–3·3
Li	3·53	0·8	0·6	≈0·5–2
K (low T)	2·29	−0·17	−0·03	≈0·01
W	6·05	3·00	1·7–2·1	≈1·2–3·4

† $\Delta S_m/k_B$ is calculated using three-quarters of the melting temperature for all elements except Na which is calculated using 361 K.

Next we must evaluate the p_y and p_z integrals. These two integrals are of the same form so that we shall focus just on the p_y integration. The limits of integration are 0 and p_{max}, where p_{max} represents a maximum kinetic energy, as yet to be determined, in a direction perpendicular to the hopping direction. The p_y integration becomes

$$\int_0^{p_{max}} \exp\left(-\frac{p_y^2}{2mk_BT}\right) dp_y = (2mk_BT)^{1/2} \int_0^{\alpha_{max}} \exp(-\alpha^2)\, d\alpha = \left(\frac{\pi mk_BT}{2}\right)^{1/2} \mathrm{erf}(\alpha_{max}). \tag{22}$$

If we require the initial trajectory of the hopping atom in the y and z directions to be such that the hopping atom avoids bumping into the neighbouring butting atoms, this places the following constraint on p_{max}:

$$p_{max} \ll (2m\mathrm{KE}_m)^{1/2}. \tag{23}$$

Also, if $p_{max}^2/2m \gtrsim 1{\cdot}5k_BT$ (a requirement virtually always met in practice), then $\mathrm{erf}(\alpha_{max}) \approx 1$ such that the p_y (and also the p_z) integration becomes

$$\int_0^{p_{max}} \exp\left(-\frac{p_y^2}{2mk_BT}\right) dp_y \approx \left(\frac{\pi mk_BT}{2}\right)^{1/2}. \tag{24}$$

Two points should be noted. First, the approximation that $\mathrm{erf}(\alpha_{max}) \approx 1$ is equivalent to letting α_{max} equal infinity. Second, a normal translational degree of freedom gives a factor of two in the numerator instead of in the denominator as found in eqn. (24); this is a consequence of having integration limits $-\infty$, ∞ for the normal translational degree case in which the particle can be travelling in either direction whereas only one direction is possible for the case considered herein.

Finally eqns. (19), (21) and (24) may be substituted into eqn. (18), which yields

$$Z_s = (1/4^{5/4}\sqrt{2})(k_BT/\mathrm{KE}_m)^{1/4}\{[(2\pi mk_BT)^{3/2}/h^3]V\} \exp(-\mathrm{KE}_m/k_BT). \tag{25}$$

Note that the term enclosed in the curly brackets is the translational partition function for a particle in a three-dimensional box, except for a factor of $e^{3/2}$ where $e = 2{\cdot}718$.

Now we must find the partition function of the hopping atom in its initial state. We assume that this is simply the partition function for a harmonic oscillator with three degrees of freedom vibrating at the Debye frequency (Lee *et al.* 1973):

$$Z_i = \{\exp(-h\nu_D/2k_BT)/[1-\exp(-h\nu_D/k_BT)]\}^3. \tag{26}$$

3.4. *The migrational entropy and enthalpy*

Now that we have obtained the partition functions for the initial and saddle-point configurations, the standard statistical thermodynamic relationships may be employed to find ΔH_m and ΔS_m (see Appendices B and C) with the final results as follows:

$$\Delta H_m = \mathrm{KE}_m + \tfrac{11}{4}k_BT - 3h\nu_D\{\tfrac{1}{2} + 1/[\exp(h\nu_D/k_BT) - 1]\} \tag{27}$$

$$\Delta S_m = \tfrac{1}{4}k_B + k_B \ln([1/4^{5/4}(2\pi)^{1/2}](k_BT/\mathrm{KE}_m)^{1/4}\{[(2\pi mek_BT)^{3/2}/h^3]V\}) - 3k_B\{(h\nu_D/k_BT)/[\exp(h\nu_D/k_BT) - 1] - \ln[1-\exp(-h\nu_D/k_BT)]\}. \tag{28}$$

Several important points need to be made with respect to these expressions. First, the enthalpy ΔH_m of migration deduced from the statistical thermodynamic derivation is that of the original BM proposal of Van Vechten with the addition of two correction terms. The first correction term $\frac{11}{4}k_BT$ arises from a $\frac{3}{2}k_BT$ contribution from the

3.3. *The partition functions*

In order to find ΔH_m and ΔS_m, we must first deduce the partition functions for the initial and saddle-point configurations, Z_i and Z_s respectively. To do this we employ the classical partition function for the saddle-point configuration:

$$Z_s = \frac{1}{h^3}\int\int_p\int\int\int_q\int \exp\left(\frac{-H_s(q_x, q_y, q_z, p_x, p_y, p_z)}{k_B T}\right) dq_x\, dq_y\, dq_z\, dp_x\, dp_y\, dp_z, \quad (16)$$

where H_s is the Hamiltonian of the hopping atom at its saddle-point configuration and the integrations are over generalized coordinates and momenta. According to the BM assumption the migration barrier is mainly kinetic in origin so that the potential energy in the Hamiltonian is neglected:

$$H_s = T_s + U_s \approx T_s = (p_x^2 + p_y^2 + p_z^2)/2m, \quad (17)$$

where T_s and U_s are the kinetic and potential energies respectively of the hopping atom at its saddle point and m is mass of the hopping atom. Substituting eqn. (17) into eqn. (16) and rearranging yields

$$Z_s = \frac{1}{h^3}\int_{q_x}\int_{q_y}\int_{q_z} dq_x\, dq_y\, dq_z \int \exp\left(-\frac{p_x^2}{2mk_B T}\right) dp_x \int \exp\left(\frac{-p_y^2}{2mk_B T}\right) dp_y \times \int \exp\left(\frac{-p_z^2}{2mk_B T}\right) dp_z. \quad (18)$$

The first triple integral is just a volume integral which represents the volume of the 'box' that the hopping atom sees when it is at its saddle-point. When the hopping atom is at its saddle-point, it is displaced partially into the volume of the vacancy that it will fill when the hop is accomplished. Additionally, new volume is created behind the hopping atom as it is displaced away from its initial equilibrium position. From these considerations, it seems reasonable that this volume is greater than that of a single vacancy but less than the volume of two isolated vacancies. If we denote this volume by V, we note that

$$\int_{q_x}\int_{q_y}\int_{q_z} dq_x\, dq_y\, dq_z = V. \quad (19)$$

Now we must evaluate the momentum integrals. Assume that the atom hops in the x direction which is collinear between the initial and final states. According to the BM hypothesis, the atom is constrained from hopping unless it possesses a kinetic energy greater than KE_m. This minimum kinetic energy for hopping corresponds to a minimum momentum p_{min} given as

$$p_{min} = (2mKE_m)^{1/2}. \quad (20)$$

Therefore the limits on the p_x integration correspond to atoms with kinetic energies greater than KE_m such that the p_x integral becomes

$$\int_{p_{min}}^{\infty} \exp\left(-\frac{p_x^2}{2mk_B T}\right) dp_x = \left(\frac{m^2(k_B T)^3}{4\,KE_m}\right)^{1/4} \exp\left(\frac{-KE_m}{k_B T}\right). \quad (21)$$

Evaluation of this integral, as indicated by the right-hand side of the above equation, is given in Appendix A.

curves in that the ordinate is total energy instead of potential energy (e.g. the difference between the initial and final state is a potential energy difference while it is a kinetic energy difference between the initial and activated complex state).

We can write an expression for the rate of the forward reaction specified by eqn. (7) by noting that it is the rate at which the activated complexes break up in their final states:

$$\text{rate} = \mathrm{d}[\mathrm{A_f}]/\mathrm{d}t = k_{\mathrm{sf}}[\mathrm{A_s}], \tag{8}$$

where the square brackets denote concentrations and k_{sf}^{-1} may now be interpreted as the average time required for an activated complex to break up. Because of the assumption of equilibrium between the initial and activated complex states we can define an equilibrium constant as follows:

$$K_{\mathrm{is}} = [\mathrm{A_s}]/[\mathrm{A_i}]. \tag{9}$$

Solving for $[\mathrm{A_s}]$ in eqn. (9) and substituting into eqn. (8) yield

$$\mathrm{d}[\mathrm{A_f}]/\mathrm{d}t = k_{\mathrm{sf}} K_{\mathrm{is}}[\mathrm{A_i}]. \tag{10}$$

From eqn. (10), we can identify the overall first-order forward rate constant k as

$$k = k_{\mathrm{sf}} K_{\mathrm{is}}. \tag{11}$$

From ordinary thermodynamics, we can relate the equilibrium constant to the Gibbs free energy ΔG_{m} of migration as

$$K_{\mathrm{is}} = \exp(-\Delta G_{\mathrm{m}}/k_{\mathrm{B}}T). \tag{12}$$

Substituting eqn. (12) into eqn. (11) yields

$$k = k_{\mathrm{sf}} \exp(-\Delta G_{\mathrm{m}}/k_{\mathrm{B}}T) = k_{\mathrm{sf}} \exp(\Delta S_{\mathrm{m}}/k_{\mathrm{B}}) \exp(-\Delta H_{\mathrm{m}}/k_{\mathrm{B}}T), \tag{13}$$

where we have expressed the Gibbs free energy of migration in terms of the corresponding enthalpy and entropy of migration. Equation (13) represents the thermodynamic formulation of absolute-rate theory for the specific case of vacancy nearest-neighbour hopping. The remainder of this section is devoted to the evaluation of k_{sf}, ΔH_{m} and ΔS_{m}.

3.2. *The attempt frequency*

First, consider k_{sf}. As noted before, the inverse of this quantity represents the average time required before an activated complex breaks up, that is before the vibration of the surrounding atoms removes the path of small potential energy between the initial and final positions. Thus it represents the lifetime of the hopping atom in its saddle-point configuration during the hop. A reasonable estimate of this lifetime would be the average zone-boundary vibrational period which we estimate as the inverse of the Debye frequency such that

$$k_{\mathrm{sf}} \approx k_{\mathrm{B}}\Theta_{\mathrm{D}}/h. \tag{14}$$

Alternatively, we could estimate k_{sf} as corresponding to the universal rate constant:

$$k_{\mathrm{sf}} \approx k_{\mathrm{B}}T/h. \tag{15}$$

Although we prefer the former estimate since it is a more accurate estimate of the vibrational fluctuations which drive solid-state diffusion, the latter estimate could also be employed and would yield essentially the same results.

Therefore, according to the model of Dobson *et al.*, the entropy of atomic migration is given by eqns. (4)–(6) while the kinetic energy contribution to the enthalpy of migration is given by eqn. (3), as proposed by Van Vechten.

§3. Statistical thermodynamics of ballistic hopping

3.1. *The rate-constant equation*

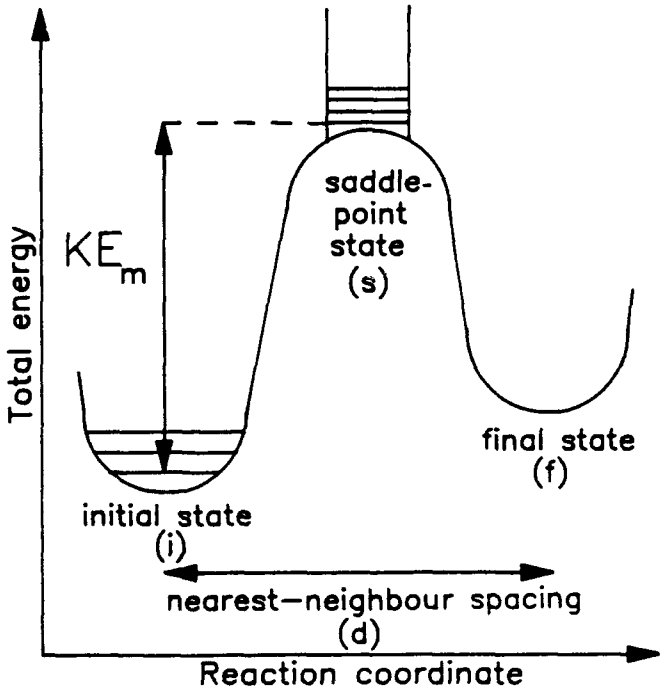
We adopt the approach of classical absolute rate theory (Beatty and Scamehorn 1977) and consider nearest-neighbour hopping as follows:

$$A_i \underset{k_{si}}{\overset{k_{is}}{\rightleftarrows}} A_s \xrightarrow{k_{sf}} A_f, \tag{7}$$

where A_i denotes the hopping atom in its initial state prior to the hop, A_s denotes the atom at its saddle-point configuration and A_f denotes the atom in its final state after the hop. k_{is} and k_{si} are the respective forward and reverse first-order rate constants for an atom going between its initial and saddle-point configuration, while k_{sf} is the first-order rate constant of the saddle-point configuration decaying into its final state. Note that we make the usual assumption of equilibrium between the initial state and its saddle-point configuration, which is also sometimes called an 'activated complex state'.

This vacancy hopping reaction is depicted in fig. 1. Note that the initial and final states need not occur at the same energy if the sites are not equivalent; in compounds, nearest-neighbour hopping leads to an asymmetry between the energies of the initial and final state because of the creation or annihilation of antisite defects concomitant with hopping. Also note that fig. 1 differs notably from typical absolute-rate theory

Fig. 1



Energy profile for vacancy nearest-neighbour hopping.

§2. The ballistic model for vacancy migration

The physical picture adopted in the BM is as follows (Van Vechten 1975). At temperatures above the Debye temperature Θ_D, the thermal vibrations of the atoms surrounding a vacancy fluctuate with a characteristic frequency given approximately by the Debye frequency $\nu_D = k_B\Theta_D/h$, where k_B is Boltzmann's constant and h is Planck's constant. Periodically these thermal fluctuations conspire to provide a path for an atom to hop into the vacancy at a cost of very little potential energy. The venue for this path is limited by the thermal motion of the surrounding atom to a period of the order of the zone-boundary phonon period of the host lattice, which is proportional to the Debye period ν_D^{-1}. For an atom to migrate successfully it must make the hop within the lifetime of the favourable venue. This requires a minimum velocity of the hopping atom to be

$$v = \nu_D d, \tag{1}$$

where d is the distance between lattice sites. This velocity corresponds to a kinetic energy

$$\mathrm{KE} = \tfrac{1}{2}mv^2, \tag{2}$$

where m is the mass of the hopping atom. Thus, since the potential energy that the atom must overcome is assumed to be quite small during this favourable venue, the kinetic energy of migration, to a good approximation, is given by

$$\mathrm{KE_m} = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m(Fd\nu_D)^2, \tag{3}$$

where F is a geometric constant equal to 0·8 for b.c.c. crystals and 0·9 for f.c.c. and h.c.p. crystals. Note that we denote this as the kinetic energy $\mathrm{KE_m}$ of migration while it was originally denoted as the enthalpy ΔH_m of migration, we anticipate the inclusion of more terms in the expression for ΔH_m, which arise from the statistical thermodynamic derivation.

In order to calculate the entropy of migration, the BM treatment was extended (Dobson *et al.* 1989). At temperatures above the Debye temperature an atom on a normal lattice site can be modelled as a simple harmonic oscillator vibrating at the Debye frequency. The entropy associated with this vibrating atom can be calculated from statistical thermodynamics (Lee, Sears and Turcotte 1973) and is given by

$$S_v/k_B = 3\{(h\nu_D/k_BT)/[\exp(h\nu_D/k_BT) - 1] - \ln[1 - \exp(-h\nu_D/k_BT)]\}, \tag{4}$$

where the factor 3 is due to the three degrees of freedom of the oscillating atom.

According to the BM, during the limited venue that the atomic migration actually occurs, that is when an atom moves through the saddle-point configuration, the hopping atom is, to a good approximation, a free particle moving ballistically. Thus for this brief period of time the vibrational modes are replaced by translational modes. We can calculate the translational entropy using the particle-in-a-box approximation (Lee *et al.* 1973):

$$S_t/k_B = \ln[(2\pi emk_BT)^{3/2}/h^3]V, \tag{5}$$

where $e = 2{\cdot}718$ and V is the volume of the box which we take to be the volume occupied by two nearest-neighbour atoms. The entropy of migration of an atom hopping into a vacancy is thus given by

$$\Delta S_m = S_t - S_v. \tag{6}$$

PHILOSOPHICAL MAGAZINE A, 1991, VOL. 63, NO. 6, 1315–1326

A statistical thermodynamic derivation of the ballistic model for vacancy migration

By J. F. WAGER

Department of Electrical and Computer Engineering,
Center for Advanced Materials Research,
Oregon State University, Corvallis, Oregon 97331, USA

[*Received 2 May 1990*† *and accepted 1 July 1990*]

ABSTRACT

A statistical thermodynamic derivation of the enthalpy and entropy associated with vacancy migration by nearest-neighbour hopping is presented. The derivation is in accordance with the ballistic model hypothesis that the migration energy is kinetic in origin for vacancy nearest-neighbour hopping. It is shown that the general form of the enthalpy and entropy of migration is qualitatively similar to, but not identical with that given previously. Except at a very high temperature or when the Debye temperature is very large, the correction of the migration enthalpy arising from the statistical thermodynamic derivation is of negligible importance. The entropy correction is, however, significant and must be employed in order to obtain agreement with experimental values of the migrational entropy for elemental metals known to diffuse by nearest-neighbour vacancy migration.

§ 1. INTRODUCTION

The ballistic model (BM) for vacancy migration was originally proposed (Van Vechten 1975) as a means of estimating the enthalpy ΔH_m of migration for vacancy nearest-neighbour hopping. The essential postulate of the BM is that the migration barrier may be envisaged as a kinetic energy barrier, instead of a potential energy barrier, as is assumed in most models. The BM model assumption has recently been employed (Dobson, Wager and Van Vechten 1989) to calculate the entropy ΔS_m of migration for vacancy nearest-neighbour hopping. In the spirit of the BM, ΔS_m was calculated as the difference between the translational entropy of the hopping atom in its saddle-point configuration and the vibrational entropy of the hopping atom in its equilibrium position prior to the hop. These two contributions to the migration entropy were calculated using standard statistical thermodynamic expressions.

The present work presents a statistical thermodynamic derivation of the BM. There are several factors which motivate this derivation. First, we wish to establish the BM on firm theoretical grounds from a classical absolute-rate theory point of view. Second, the statistical thermodynamic derivation leads to new insight into the physics inherent in the BM assumptions. Third, explicit expressions for ΔH_m and ΔS_m are deduced which are then compared with those given previously. It is shown that the statistical thermodynamic expressions are similar to, but not identical with, expressions given previously.

† Received in final form 14 June 1990.

THERMODYNAMICS AND KINETICS OF VACANCY SELF-COMPENSATION IN WIDE BAND GAP SEMICONDUCTORS

J.F. Wager

Department of Electrical and Computer Engineering
Center for Advanced Materials Research
Oregon State University
Corvallis, OR 97331-3211

ABSTRACT

Thermodynamic and kinetic aspects of anion vacancy self-compensation of p-type layers in three compound semiconductors, GaN, ZnSe, and SiC, are considered. Thermodynamic considerations indicate that SiC p-type layers are thermodynamically stable, whereas GaN and ZnSe p-type layers are thermodynamically unstable, with respect to anion vacancy self-compensation. Hence, if GaN and ZnSe compound semiconductor devices with p-type layers are to exhibit long-term stability, it is important that kinetic barriers be established which preclude the supply of anion vacancies to the p-type layer. Kinetic strategies for minimizing anion vacancy self-compensation include growing p-layers free of grain boundaries and dislocations, placing the p-layers remote from surfaces, and embedding the p-layer in heavily n-doped regions.

"Submitted to Phil. Mag. A"

1. INTRODUCTION

There is a resurgence of interest in wide band gap semiconductors (Bhargava 1989, Kukimoto 1989). This is partially motivated by recognition of the potential of these semiconductors for applications such as visible displays, visible/ultraviolet optical sources or detectors, and high-temperature electronic devices. However, recognition of the technological potential of these materials is not the primary reason for this recent increase in research activity. Rather, it is primarily driven by the emergence of new process technologies, the utilization of non-traditional dopants, and the adoption of new doping strategies which have led to the attainment of low-resistivity, n- and p-type doping in wide band gap semiconductors.

Until recently, it has been very difficult to obtain low-resistivity, n- and p-type layers for many wide band gap semiconductors. Certain semiconductors (e.g., ZnSe, ZnS, CdS, GaN) exhibit a propensity for only low-resistivity, n-type doping whereas other semiconductors (e.g., ZnTe) tend to give rise to exclusively low-resistivity, p-type doping.

This inability to achieve bipolar doping in many wide band gap semiconductors is attributed (Aven and Devine 1973, Park and Shin 1977) to the low solubility (Kröger 1964, Neumark 1989) or the high ionization energy of dopants, to impurity compensation (Henry, Nassau, and Shiever 1971), or to one of several kinds of self-compensation (Mandel 1964, Neumark 1980, Van Vechten 1980, Chadi and Chang 1989).

The first purpose of the work described herein is to explore the thermodynamics of a specific Fermi-level-driven vacancy self-compensation mechanism for three wide band gap semiconductors, GaN, ZnSe, and SiC. According to these estimates, p-type doping in GaN and ZnSe is thermodynamically unstable with respect to anion vacancy self-compensation whereas p-type doping is found to be thermodynamically stable for SiC.

Thermodynamics is not the entire story, however, since thermodynamically unstable material systems can be rendered metastable if an appropriate kinetic barrier exists. An obvious example of such

a situation is carbon in the form of diamond which is unstable (actually, metastable) with respect to the thermodynamically more stable graphite state; it is the existence of a large kinetic barrier which impedes the spontaneous transformation of diamond to graphite. Thus, the second purpose of the present paper is to explore the anion vacancy self-compensation kinetics of GaN and ZnSe in order to estimate the potential stability of devices fabricated using these wide band gap semiconductors whose p-type doping is unstable with respect to anion vacancy self-compensation. Our assessment of the self-compensation kinetics suggests that it will be extremely difficult to manufacture stable GaN or ZnSe devices with low-resistivity p-type layers unless these devices are designed in a manner such that anion vacancy diffusion to the p-type layers can be suppressed.

The Fermi-level-driven vacancy self-compensation mechanism invoked in our analysis is not presumed to universally account for all problems associated with bipolar doping in wide band gap semiconductors. Rather, it is believed that the compensation of certain wide band gap semiconductors may be attributed to this mechanism. It is also believed that this mechanism may play an important role in establishing the long-term stability of devices fabricated using certain wide band gap semiconductors. Finally, it is hoped that the basic thermodynamic/kinetic approach utilized in this paper may provide a conceptual framework from which to evaluate other instability mechanisms and, perhaps, to lead to the implementation of new strategies for obtaining larger kinetic barriers, thus, resulting in improved device stability.

2. THERMODYNAMICS OF VACANCY SELF-COMPENSATION

The analytical formulation employed herein to account for vacancy self-compensation relies only on knowing the enthalpy of formation, ΔH_f, of the appropriate vacancy. ΔH_f of an isolated, ionized vacancy is envisaged (Van Vechten 1975a, Van Vechten 1980, Dobson and Wager 1989, Wager 1991) as the sum of the enthalpy of formation of a neutral vacancy and an electronic contribution which is related to the ionization energy of the vacancy;

$$\Delta H_f(V^z) = \Delta H_f(V^x) + \Delta H_e(V^z) \quad (1)$$

where the superscript z denotes the charge state of the vacancy and x denotes the vacancy as neutral. The thermodynamic stability of a semiconductor with respect to vacancy self-compensation is simply assessed by the condition

$$\Delta H_f(V^z) = 0 \quad (2)$$

which establishes the critical Fermi-level position, denoted E_F^{crit}, beyond which it is more energetically favorable to create new, self-compensating vacancies rather than to modulate the Fermi-level closer to the appropriate band edge. It is most convenient to illustrate this approach with specific examples.

Consider the case of p-type, acceptor doping in GaN. A bulk energy band diagram relevant to such a situation is indicated in Fig. 1. The nitrogen vacancy, V_N, a single donor with an ionization energy of approximately 0.05 eV (Madelung 1982) with respect to the conduction band minimum, E_C, is relevant for the p-type doping case under consideration. The 0/+ convention employed in Fig. 1 refers to the ionization state of the level when the Fermi-level is sufficiently far above or below the ionization level. Note that the energy separation shown in Fig. 1 is not drawn to scale. In Fig. 1, the valence band maximum, E_V, is taken as the energy reference. The band gap indicated in Fig. 1 is that of room temperature. In our analysis the relevant temperature is that of device operation, approximated by room temperature, and not the semiconductor growth temperature since our primary concern is with the long-term stability of compound semiconductor devices.

The expression for the enthalpy of formation of V_N corresponding to Eqn. 1 is (Dobson and Wager 1989, Wager 1991)

$$\Delta H_f(V_N^*) = \Delta H_f(V_N^x) + (E_F - E_D) . \quad (3)$$

E_F^{crit} is found by setting $\Delta H_f(V_N^+) = 0$ in Eqn. (3) to yield

$$E_F^{crit} = E_D - \Delta H_f(V_N^x) . \quad (4)$$

From Fig. 1, E_D = 3.39 eV and according to Van Vechten's microscopic cavity model (MCM) (Van Vechten 1975a, Van Vechten 1980), $\Delta H_f(V_N^x)$ = 1.99 eV so that Eqn. (4) gives

$$E_F^{crit}(\text{GaN, p-type doping}) = 1.4 \text{ eV} . \tag{5}$$

Equation (5) is the condition for the onset of self-compensation in GaN by nitrogen vacancy creation; Eqn. (5) implies that when the acceptor doping density is large enough that the Fermi-level is less than 1.4 eV above E_V, it is energetically more favorable to create nitrogen vacancies which compensate any additional acceptors. Alternatively, Eqn. (5) is interpreted as the energy at which the Fermi-level is pinned at high acceptor doping concentrations, assuming that equilibrium prevails. It is crucially important to realize that Eqn. (5) does not mean that E_F in GaN cannot be positioned closer to E_V than 1.4 eV; rather, the message of Eqn. (5) is that if E_F is less than 1.4 eV, our estimate indicates that this situation is not thermodynamically stable with respect to V_N self-compensation. Thus, our analysis precludes the attainment of equilibrium, low-resistivity, p-type layers in GaN because of the unfavorable energetics of anion vacancy self-compensation.

Now consider the case of p-type, acceptor doping in ZnSe, whose energy band diagram is illustrated in Fig. 2. The selenium vacancy, V_{Se}, is a double donor with ionization energies of approximately 0.02 eV and 0.3 eV with respect to E_C (Shirakawa and Kukimoto 1980, Ito and Okada 1985). Therefore, the ZnSe equivalent of Eqn. (1) is (Dobson and Wager 1989, Wager 1991)

$$\Delta H_f\left(V_{Se}^{2+}\right) = \Delta H_f\left(V_{Se}^{x}\right) + \left(E_F - E_{D1}\right) + \left(E_F - E_{D2}\right) . \tag{6}$$

E_F^{crit} is found by setting $\Delta H_f(V_{Se}^{2+}) = 0$ which results in

$$E_F^{crit} = \frac{1}{2}\left[E_{D1} + E_{D2} - \Delta H_f\left(V_{Se}^{x}\right)\right] . \tag{7}$$

Using E_{D1} = 2.68 eV, E_{D2} = 2.4 eV from Fig. 2 and $\Delta H_f(V_{Se}^x)$ = 2.56 eV from the MCM estimate (Van Vechten 1980) yields

$$E_F^{crit}(\text{ZnSe, p-type doping}) = 1.3 \text{ eV} . \quad (8)$$

Similar to GaN, our vacancy self-compensation analysis indicates that it is not possible to obtain low-resistivity, p-type ZnSe layers under equilibrium conditions.

Finally, consider the thermodynamics of vacancy self-compensation for cubic SiC under p-type, acceptor doping. Figure 3 illustrates the bulk energy band diagram. The carbon vacancy, V_C, is taken to be a simple donor (More, Ryu, Carter, Bentley, and Davis 1985, Li and Lin-Chung 1987, Talwar and Feng 1991) with an ionization energy E_D = 1.66 eV (Talwar and Feng 1991). It should be noted that E_D has also been estimated as 0.54 eV (Li and Lin-Chung 1987). Even though there is a large disparity of these estimates of E_D such that the V_C donor ionization energy should be considered uncertain, our conclusion regarding the vacancy self-compensation stability of SiC with respect to p-type doping is independent of which estimate of E_D is used.

The SiC equivalent of Eqn. (1) is

$$\Delta H_f(V_C^+) = \Delta H_f(V_C^x) + (E_F - E_D) . \quad (9)$$

Setting $\Delta H_f(V_C^+) = 0$ gives

$$E_F^{crit} = E_D - \Delta H_f(V_C^x) . \quad (10)$$

The MCM estimate for $\Delta H_f(V_C^x)$ is 2.25 eV (Van Vechten 1980) so that

$$E_F^{crit} = -0.59\text{eV} . \quad (11)$$

This large, negative estimate for E_F^{crit} implies that, in contrast to GaN and ZnSe, SiC is stable with respect to anion vacancy self-compensation under strong p-type doping. Thus, according to the present analysis it is possible to obtain equilibrium, low-resistivity, p-type SiC layers without being constrained by the energetics of anion vacancy self-compensation.

In summary, thermodynamic arguments are presented which indicate GaN and ZnSe to be thermodynamically unstable, whereas SiC is found to be stable, with respect to anion vacancy self-

compensation under strong p-type doping. It should be noted that a similar analysis under strong n-type doping indicates that GaN, ZnSe, and SiC are all stable with respect to cation vacancy self-compensation.

3. KINETICS OF SELF-COMPENSATION IN GaN AND ZnSe

The conclusion of the previous analysis is that p-type doping in GaN and ZnSe is not thermodynamically stable with respect to anion vacancy self-compensation. Hence, the challenge facing the fabrication of stable GaN and ZnSe devices which employ low-resistivity, p-type layers is to devise strategies for obtaining suitable kinetic barriers which render these layers adequately stable (i.e., stable in a kinetic sense, not a thermodynamic sense).

Our kinetic assessment begins by assuming the existence of a low-resistivity, p-type GaN or ZnSe layer which is incorporated into a device structure. The growth of such a p-type layer will most likely be achieved through the use of advanced crystal growth technologies, novel dopants or doping strategies, and the precise control of stoichiometry and impurities (Bhargava 1989, Kukimoto 1989). Although the details of how such a layer is actually grown is of no further interest for our present purposes, it is important to note that such a low-resistivity, p-type layer is not in thermodynamic equilibrium, according to our previous thermodynamic analysis. Thus, anion vacancy self-compensation is a natural degradation mechanism for such a device.

We consider three means by which anion vacancies can be supplied in order that self-compensation occur in the p-type layer:

1. Anion vacancies are generated spontaneously within the p-type layer if vacancy-nucleation sites such as surfaces, grain boundaries, or dislocations are available to nucleate such vacancies.
2. Anion vacancies outside the p-type layer diffuse into the p-type layer.
3. Anion vacancies are generated within the p-type layer by Frenkel-pair generation.

We now explore in more detail the kinetics associated with supplying anion vacancies for self-compensation.

First, consider the case of a low-resistivity, p-type layer which is in direct contact with a surface or which possesses a high concentration of dislocations or grain boundaries so that anion vacancies are readily nucleated. Our previous thermodynamic analysis suggests that anion vacancies generate spontaneously at nucleation sites since there is no thermodynamic barrier precluding their formation. Therefore, self-compensation of the p-type layer depends on the rate at which vacancies diffuse from vacancy nucleation sites throughout the p-type layer. The activation energy for such a process is equal to the enthalpy of migration, ΔH_m, for diffusion which we estimate to be approximately 0.9 eV and 1.8 eV for anion vacancy migration in GaN and ZnSe, respectively. These ΔH_m estimates are based on the ballistic model for atomic migration (Van Vechten 1975b, Van Vechten 1980) assuming that anion vacancy migration in GaN and ZnSe occurs by second-nearest-neighbor hopping (2nnh).

Consequently, we estimate the degradation activation energies to be 0.9 eV and 1.8 eV for GaN and ZnSe devices, respectively, if degradation occurs by anion vacancy nucleation within the p-type layer. These degradation activation energies are significantly smaller than the relevant semiconductor band gap and it is expected that they are of an inadequate magnitude to insure the stability of such compound semiconductor devices. Thus, the first conclusion arising from this kinetic analysis is that low-resistivity, p-type layers should be located remote from surfaces and should be as free as possible from dislocations, grain boundaries, or any other vacancy-nucleation sites.

The second way of accomplishing anion vacancy self-compensation is via the diffusion of anion vacancies from device regions outside the low-resistivity, p-type layer. Suppose that our low-resistivity, p-type layer is completely embedded within an n-type material of identical composition to the p-layer, thus forming a homojunction. ΔH_f of the charged vacancies depends on the position of E_F. Recall that ΔH_f for a GaN anion vacancy is given by

$$\Delta H_f(V_N^{\cdot}) = \Delta H_f(V_N^{x}) + (E_F - E_D) \tag{12}$$

whereas for ZnSe it is given by

$$\Delta H_f(V_{Se}^{2\cdot}) = \Delta H_f(V_{Se}^{x}) + (E_F - E_{D1}) + (E_f - E_{D2}) . \tag{13}$$

From Eqns. (12) and (13) it is evident that ΔH_f for anion vacancies increases monotonically as E_F approaches E_C until the donor is neutral. The maximum ΔH_f occurs when E_F crosses the uppermost donor ionization level, rendering the anion vacancy neutral. According to the MCM estimates (Van Vechten 1980), this maximum corresponds to $\Delta H_f(V_N^x) = 1.99$ eV and $\Delta H_f(V_{Se}^x) = 2.56$ eV for GaN and ZnSe, respectively. Thus, anion vacancy formation is thermodynamically more difficult in the n-type embedding region than in the low-resistivity, p-type layer. Assuming that the n-type region is grown relatively vacancy-free, new vacancies must be generated at vacancy-nucleation sites within the n-type layer and subsequently diffuse through the n-type region to the low-resistivity, p-type layer to accomplish anion vacancy self-compensation. The activation energy, ΔH_a, for this vacancy-generation, diffusion process is given by

$$\Delta H_a = \Delta H_f(V^x) + \Delta H_m^{2nnh} . \tag{14}$$

If the n-type region is heavily doped to insure that the anion vacancies are neutral, ΔH_a is approximately 2.9 eV and 4.3 eV for GaN and ZnSe, respectively. ΔH_a corresponds to the activation energy of device degradation by anion vacancy self-compensation in which anion vacancies are supplied by remote generation and diffusion. Note that this activation is maximal for n^+ doping of the embedding region and is correspondingly reduced as the embedding region doping concentration is decreased.

It is not clear whether these activation energies, 2.9 eV and 4.3 eV, are of sufficient magnitude to guarantee the stability of GaN and ZnSe compound semiconductor devices which contain low-resistivity, p-type layers. As a rule of thumb, we believe that ΔH_a should be greater than the semiconductor band gap since this amount of energy is readily available from recombination of carriers across the band gap.

From this point of view, ZnSe devices appear to hold more promise in terms of long-term stability than GaN devices. In any event, two strategies emerge from this kinetic analysis:

1. The p-type layer should be embedded within an n-type region of high doping concentration in order to provide a large activation energy for vacancy-generation, diffusion.
2. The n-type embedding region should have as large dimensions as possible to impede anion vacancy diffusion to the low-resistivity, p-type layer.

Now consider Frenkel-pair generation, the third way in which anion vacancies may be supplied for self-compensation of the p-type layer. Frenkel-pair generation corresponds to the creation of a self-interstitial-vacancy pair, N_i-V_n and Se_i-V_{Se} for GaN or ZnSe, respectively. If we employ Van Vechten's method (Van Vechten 1980) for estimating ΔH_f of an anion self-interstitial, we obtain $\Delta H_f(N_i^x$ in GaN) = 27 eV and $\Delta H_f(Se_i^x$ in ZnSe) = 18 eV. These estimates seem unrealistically large and certain assumptions underlying this approach have been questioned (Morgan-Pond and Raghavan 1985). However, *ab initio* pseudo-atomic orbital calculations (Jansen and Sankey 1989) also yield a large estimate of about 12 eV for ΔH_f of a neutral anion Frenkel-pair in ZnSe. Therefore, we conclude from these estimates that anion Frenkel-pair generation is not likely to be an important mechanism for supplying vacancies for the self-compensation of p-type layers in GaN and ZnSe.

In summary, because p-type doping of GaN and ZnSe is not thermodynamically stable with respect to anion vacancy self-compensation, such p-type layers should be located remote from surfaces or other anion vacancy nucleation sites. GaN and ZnSe device stability can also be improved if p-type layers are embedded within n-type layers. Frenkel-pair generation is not believed to be important in determining the stability of GaN and ZnSe devices.

4. CONCLUSIONS

The thermodynamics and kinetics of anion vacancy self-compensation is investigated for the compound semiconductors GaN, ZnSe, and SiC. SiC p-type layers are found to be thermodynamically stable with respect to anion vacancy self-compensation whereas GaN and ZnSe p-type layers are thermodynamically unstable. Since anion vacancy self-compensation is an obvious device degradation mechanism, the best means of insuring long-term stability is to employ stable semiconductors such as SiC. If semiconductors which are unstable with respect to anion vacancy self-compensation are to be utilized, certain strategies to improve device stability should be adopted. These strategies consist of growing the p-type semiconductor free of vacancy nucleation sites such as grain boundaries and dislocations, placing the p-type layers far from surfaces or other vacancy nucleation sites which exist in nearby layers, and embedding the p-type layers in heavily doped n-type layers.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Air Force Office of Scientific Research under contract AFOSR 89-0309.

REFERENCES

Aven, M. and Devine, J.Z., 1973, *J. Luminesc.*, *7*, 195.

Bhargava, R.N., 1989, *Growth and Optical Properties of Wide-Gap II-VI Low-Dimensional Semiconductors*, edited by T.C. McGill, C.M. Sotomayor Torres, and W. Gebhardt (New York:Plenum), p. 1.

Chadi, D.J. and Chang, K.J., 1989, *Appl. Phys. Lett.*, *55*, 575.

Das, K., Kong, H.S., Petit, J.B., Bumgarner, J.W., and Davis, R.F., 1990, *J. Electrochem. Soc.*, *137*, 1598.

Dobson, T.W. and Wager, J.F., 1989, *J. Appl. Phys.*, *66*, 1997.

Henry, C.H., Nassau, K., and Shiever, J.W., 1971, *Phys. Rev. B*, *4*, 2453.

Ido, T. and Okada, M., 1985, *J. Crystal Growth*, *72*, 170.

Jansen, R.W. and Sankey, O.F., 1989, *Phys. Rev. B*, *39*, 3192.

Kröger, F.A., 1974, *The Chemistry of Imperfect Crystals*, 2nd edition, (Amsterdam:North-Holland).

Kukimoto, H., 1989, *Growth and Optical Properties of Wide-Gap II-VI Low-Dimensional Semiconductors*, edited by T.C. McGill, C.M. Sotomayor Torres, and W. Gebhardt (New York:Plenum), p. 119.

Li, Y. and Lin-Chung, P.J., 1987, *Phys. Rev. B*, *36*, 1130.

Madelung, O., editor, 1989, *Landolt-Börstein*, vol. 22 (Berlin:Springer-Verlag), p. 502.

Mandel, G., 1964, *Phys. Rev.*, *134*, A1073.

More, K.L., Ryu, J., Carter, C.H., Bentley, J., and Davis, R.F., 1985, *Cryst. Latt. Def. and Amor. Mat.*, *12*, 243.

Morgan-Pond, C.G. and Raghavan, R., 1985, *Phys. Rev. B*, *31*, 6616.

Neumark, G.F., 1980, *J. Appl. Phys.*, *51*, 3383.

Neumark, G.F., 1989, *Phys. Rev. Lett.*, *62*, 1800.

Park, Y.S. and Shin, B.K., 1977, *Electroluminescence*, edited by J.I. Pankove (Berlin:Springer-Verlag), p. 133.

Shirakawa, Y. and Kukimoto, H., 1980, *Solid State Commun.*, *34*, 359.

Talwar, D.N. and Feng, Z.C., 1991, *Phys. Rev. B*, *44*, 3191.

Van Vechten, J.A., 1975a, *J. Electrochem. Soc.*, *122*, 419.

Van Vechten, J.A., 1975b, *Phys. Rev. B*, *12*, 1247.

Van Vechten, J.A., 1980, *Handbook on Semiconductors*, vol. 3, edited by S.P. Keller (Amsterdam:North-Holland), chap. 1.

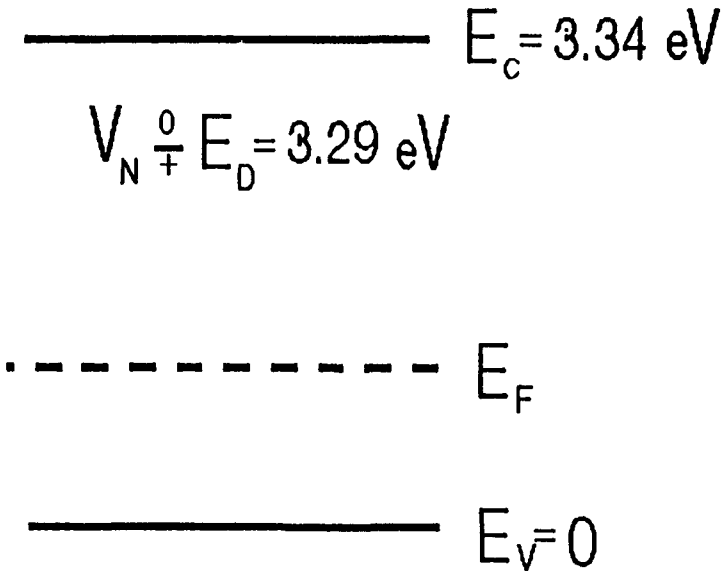
Wager, J.F., 1991, *J. Appl. Phys.*, *69*, 3022.

FIGURE CAPTIONS

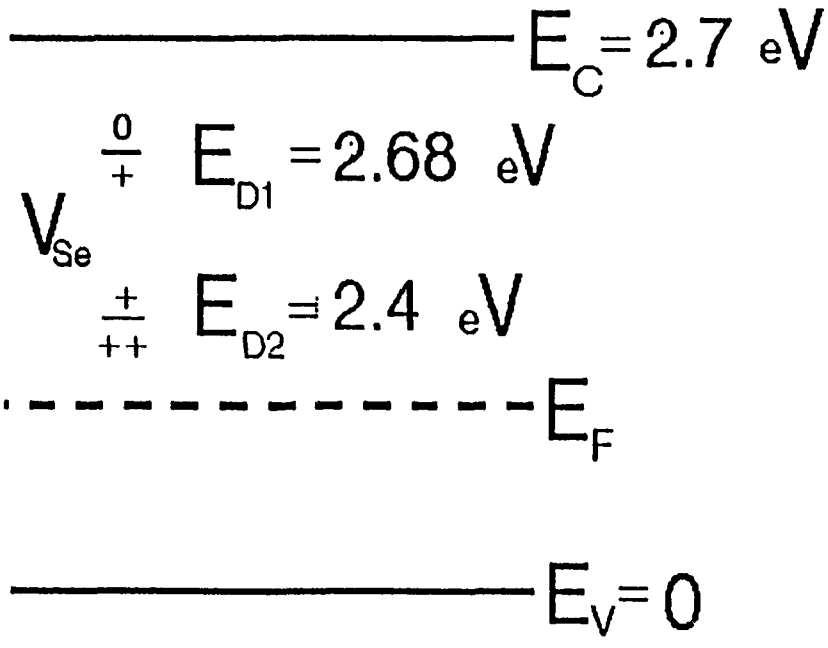
Figure 1. Bulk energy band diagram of GaN.

Figure 2. Bulk energy band diagram of ZnSe.

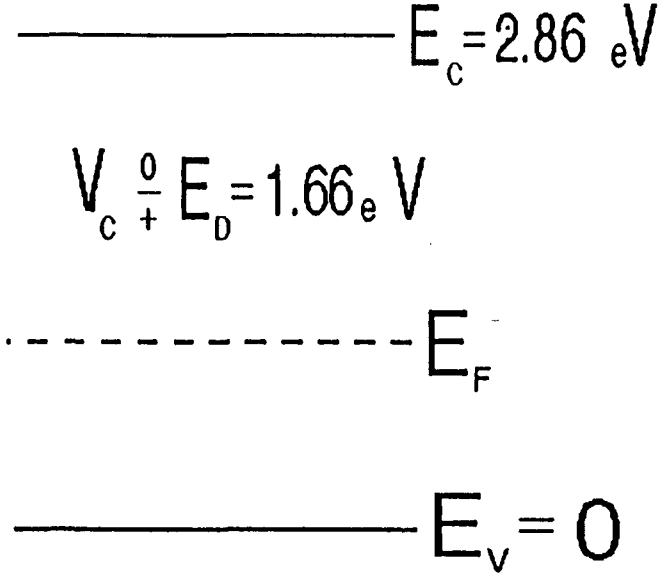
Figure 3. Bulk energy band diagram of SiC.



WAGER FIGURE 1



WAGER FIGURE 2



WAGER FIGURE 3